Rôle des familles géniques YABBY et ARF dans la mise en place du carpelle au cours de l'évolution.

Cédric Finet

N° d'ordre : 476 **N° attribué par la bibliothèque : 07ENSL0 476**

THÈSE

en vue d'obtenir le grade de

Docteur de l'Université de Lyon – École Normale Supérieure de Lyon

Spécialité : Sciences de la Vie
Laboratoire de Reproduction et Développement des Plantes
École doctorale : Biologie Moléculaire, Intégrative et Cellulaire

présentée et soutenue publiquement le 29/09/2008

par Monsieur Cédric FINET

Rôle des familles géniques YABBY et ARF dans la mise en place du carpelle au cours de l'évolution

---

Directeur de thèse : Monsieur Charlie SCUTT

Après avis de : Madame Sophie NADOT, *Membre/Rapporteur*
Monsieur Marc ROBINSON-RECHAVI, *Membre/Rapporteur*

Devant la Commission d'examen formée de :

Monsieur Guillaume BALAVOINE, *Membre*
Madame Sophie NADOT, *Membre/Rapporteur*
Monsieur Marc ROBINSON-RECHAVI, *Membre/Rapporteur*
Monsieur Charlie SCUTT, *Membre*

N° d'ordre : 476 **N° attribué par la bibliothèque : 07ENSL0 476**

THÈSE

en vue d'obtenir le grade de

Docteur de l'Université de Lyon – École Normale Supérieure de Lyon

Spécialité : Sciences de la Vie
Laboratoire de Reproduction et Développement des Plantes
École doctorale : Biologie Moléculaire, Intégrative et Cellulaire

présentée et soutenue publiquement le 29/09/2008

par Monsieur Cédric FINET

Rôle des familles géniques YABBY et ARF dans la mise en place du carpelle au cours de l'évolution

---

Directeur de thèse : Monsieur Charlie SCUTT

Après avis de : Madame Sophie NADOT, *Membre/Rapporteur*
Monsieur Marc ROBINSON-RECHAVI, *Membre/Rapporteur*

Devant la Commission d'examen formée de :

Monsieur Guillaume BALAVOINE, *Membre*
Madame Sophie NADOT, *Membre/Rapporteur*
Monsieur Marc ROBINSON-RECHAVI, *Membre/Rapporteur*
Monsieur Charlie SCUTT, *Membre*

Résumé

L'apparition de la fleur a nécessité trois événements majeurs: (i) le regroupement des organes reproducteurs mâles et femelles sur un même axe, (ii) l'internalisation des ovules au sein d'un organe: le carpelle, (iii) la mise en place de pièces stériles en périphérie des organes reproducteurs. Ce travail de thèse consiste en l'identification d'événements moléculaires à l'origine du carpelle. Les gènes *ARF3* et *ARF4* jouent un rôle clé dans le développement du carpelle chez l'espèce modèle *Arabidopsis thaliana*. La reconstruction de l'histoire évolutive de ces gènes a permis de montrer qu'ils résultaient d'une duplication dans le lignage menant aux angiospermes. Leurs structures indiquent qu'ils ont évolué par perte de certains domaines protéiques, ceci de manière indépendante dans les lignages ARF3 ou ARF4. Ces changements dans la partie codante constituent un mode d'évolution généralisable à l'ensemble des embryophytes. La famille génique YABBY intervient dans l'établissement de la polarité adaxiale-abaxiale des organes latéraux. En particulier, les gènes *CRC* et *INO* constituent respectivement deux marqueurs moléculaires du carpelle et de l'ovule. L'étude préliminaire de cette famille semble indiquer l'absence du gène *CRC* chez les gymnospermes, suggérant que l'apparition de *CRC* aurait été un pré-requis pour l'origine évolutive du carpelle.

Abstract

The origin of the flower required three main steps: (i) the coming together of male and female reproductive organs of a single axis, (ii) the closure of a fertile bract to form the carpel, (iii) the development of sterile organs to form the perianth. The present work describes the identification of molecular events that could have played a role in the evolutionary appearance of carpel. The genes *ARF3* and *ARF4* are involved in carpel development in the model species *Arabidopsis thaliana*. The reconstruction of the evolutionary history of these two genes suggests they result from a preangiosperm gene duplication event. Moreover, these genes may have predominantly evolved by independent loss of functional regions in the ARF3 or ARF4 lineages. Such changes to coding sequences seem to have occurred frequently in the evolution of the ARF gene family in the land plants. YABBY genes represent a second class of key genes in the establishment of the adaxial-abaxial polarity of lateral organs. In particular, *CRC* and *INO* represent excellent molecular markers of the carpel and the ovule, respectively. The preliminary study of the YABBY gene family indicates that *CRC* might be absent in the sister group of the extant flowering plants, or gymnosperms. This data suggests a link between the origin of the gene *CRC* and the appearance of the carpel in angiosperms.

AVANT-PROPOS

Mes premiers remerciements sont pour les quelques téméraires qui ont accepté de lire ma prose en entier (du moins je l'espère) : Sophie Nadot, Marc Robinson-Rechavi, Guillaume Balavoine et Charlie Scutt.

Je tiens également à remercier Christian et Charlie qui m'ont donné la chance de connaître le RDP qui est définitivement un endroit fabuleux, que ce soit sur le plan scientifique ou humain. Maintenant, c'est au tour de Jan de veiller à ce que le labo conserve cette même ambiance ;)

Je me souviens…

Je me souviens de Claudia et de Pierrot dansant entre les paillasses.
Je me souviens de ces soirées au labo, de ces discussions ésotériques sur la théorie de l'évolution avec mes compères Mich et Séb.
Je me souviens de Gégé, du chiendent et de la Saint Glinglin.
Je me souviens du couloir Sud-est, Christophe tenant sa nièce dans ses bras, heureux comme... un futur papa.
Je me souviens des chocolats glissés dans la boîte aux lettres pour un rendez-vous manqué.
Je me souviens de Pierrot l'anachorète.
Je me souviens de Guiguaill se promenant au parc.
Je me souviens de Sophie, de ses boulimies de framboises, de myrtilles.
Je me souviens d'Olivier et de son garde du corps.
Je me souviens des canettes de Coca-Cola sur le bureau de Vincent.
Je me souviens d'Hervé et de sa culture générale inégalée au labo.
Je me souviens de Miss Martinique.
Je me souviens de Peter brandissant une épée chez Cyril.
Je me souviens d'Annick à 70%.
Je me souviens la fois où j'ai battu Yvon aux colons.
Je me souviens de Magalie C. et de ses lapins.
Je me souviens d'Isa et de ses rendez-vous au Hilton.
Je me souviens du rire de Sandrine.
Je me souviens de Fanfan et de ses huiles essentielles.
Je me souviens de Pradeep, ennemi du sucre, ami des goûters.
Je me souviens de Thierry enRHUMé.
Je me souviens des deux grands-pères Aillevonne et Mik, ah ouais c'était le bon temps.
Je me souviens de Frédérique et de sa trottinette.
Je me souviens de Madame Rogowsky.
Je me souviens du papa d'Angèle, de Roxanne… et d'Émile.
Je me souviens de "L'affaire Marianne Schaedel".
Je me souviens de Nathalie à qui je décerne le titre de Mère Courage.
Je me souviens de Mik râlant.
Je me souviens d'Alexis et de son efficacité pour lutter contre les fumeurs.
Je me souviens de la dionée d'Olivier.

Je me souviens du feuilleton “Sylvie et le chien méchant”.

Je me souviens de Vanessa après lui avoir annoncé la réussite de Catherine.

Je me souviens de Davide passant des coups de fil en italien, les pieds sur le bureau de Peter.

Je me souviens de Nico et de sa joie de vivre.

Je me souviens de Magalie U. et de sa gentillesse à toute épreuve.

Je me souviens de la tante Laetitia et de la scribouillarde.

Je me souviens d’Anne-marie et du cadeau velouté qu’elle me fit.

Je me souviens d’Adeline et de ses frites à moitié mangées.

Je me souviens d’Edwige et de nos fous rires.

Je me souviens du pot du groupe SiCE.

Je me souviens de Teva vivant à l’ENS pendant un mois.

Je me souviens de Patrice et de son humour caustique.

Je me souviens de Nelu et de sa clarté obscure.

Je me souviens de Véro et de ses victoires RACE-plendissantes.

Je me souviens de Florian nu dans la serre.

Je me souviens de Marion et de son parfum à la rose.

Je me souviens de Marty et de ses jupes courtes.

Je me souviens de Christine et de sa fraîcheur.

Je me souviens de Catherine qui m’invite à manger chez elle un midi, ou pas.

Je me souviens d’une certaine voisine dont j’ai oublié le prénom.

Je me souviens de Mich et de sa vision à long terme.

Je me souviens du déménagement de Séb, de Magalie, de Mich, de Nicolas, d’Olivier.

Je me souviens du renard orange de Séb comme preuve irréfutable de la non-existence de la sélection naturelle.

Je me souviens des goûters entre filles.

Je me souviens d’Aurélie peignant consciencieusement la pendule du coin café.

Je me souviens de Valérie et de l’épopée de Pimousse et de Croquette.

Je me souviens de Priscilla et Isabelle riant sur le compte de l’intersaison.

Je me souviens de Guiguaill et de la camionnette à pizzas.

Je me souviens d’Aillevonne et de ses impact factors.

Je me souviens de ces instants volés au RDP…

ABRÉVIATIONS

ADN	acide désoxyribonucléique
ADNc	ADN complémentaire
ARN	acide ribonucléique
ARNm	ARN messager
DBD	DNA-binding domain
EMS	ethyl methane sulfonate
EST	expressed sequence tag
GUS	β-glucuronidase
Ma	millions d'années
NPA	N-1-naphthylphtalamic acid
nt	nucléotides
pb	paires de bases
PCR	polymerase chain reaction
RACE	rapid amplification of cDNA ends
RT	reverse transcription
WT	wild type (sauvage)

CONVENTIONS

GÈNE

PROTÉINE

mutant

AVANT-PROPOS

La problématique majeure de ce travail de thèse consiste en l'origine évolutive des plantes à fleurs. Ce travail vise à identifier des événements moléculaires qui auraient pu contribuer à l'apparition des plantes à fleurs sur la Terre. Pour un projet spécifique (chapitre 3), l'analyse a permis de tirer des conclusions sur l'ensemble des plantes terrestres, et a donc par conséquent outrepassé la question de l'apparition des plantes à fleurs. Au final, ce travail apporte des éléments de réponse à la question en utilisant essentiellement une approche de génétique évolutive du développement du carpelle (ou "évo-dévo").

Comme l'indique clairement le sommaire, le manuscrit de thèse se divise en différentes parties :

- une introduction, ou travail de recherche bibliographique (chapitre 1),
- une présentation de la problématique et de la méthodologie suivie (chapitre 1),
- une présentation des principaux résultats scientifiques, sous la forme de manuscrits rédigés en anglais et qui seront envoyés à des journaux scientifiques à plus ou moins court terme (chapitres 2, 3, 4 et 5),
- une discussion générale sur l'ensemble des résultats obtenus (chapitre 6).

L'introduction se divise en quatre parties. Dans une première partie, je fais un certain nombre de rappels concernant la biologie des principaux groupes de plantes terrestres. Tout d'abord, je présente les différents embranchements dans un contexte phylogénétique. Puis, je récapitule brièvement les cycles biologiques ainsi que les principaux caractères morphologiques de chacun de ces groupes. Le but de cette partie est double : (i) souligner le fait que la fleur (en tant qu'entité reproductrice) est le résultat d'un long processus évolutif, (ii) présenter des groupes peu connus comme les lycophytes pour lesquelles sont disponibles des données génomiques que j'ai étudié pendant ma thèse (chapitre 3).

Dans une deuxième partie, je fais un bilan des différentes données que nous avons à l'heure actuelle pour essayer de répondre à la question de l'origine des plantes à fleurs. Comme vous le verrez, les données sont à la fois d'ordre paléobotanique (registre fossile), phylogénétique (basée sur des espèces actuelles), et moléculaires.

Dans une troisième partie, je dresse un bilan des connaissances acquises en génétique du développement du carpelle chez l'espèce modèle *Arabidopsis thaliana*, ainsi que des principales études de génétique évolutive du développement du carpelle qui en ont découlées.

INTRODUCTION

Chapitre 1 : INTRODUCTION

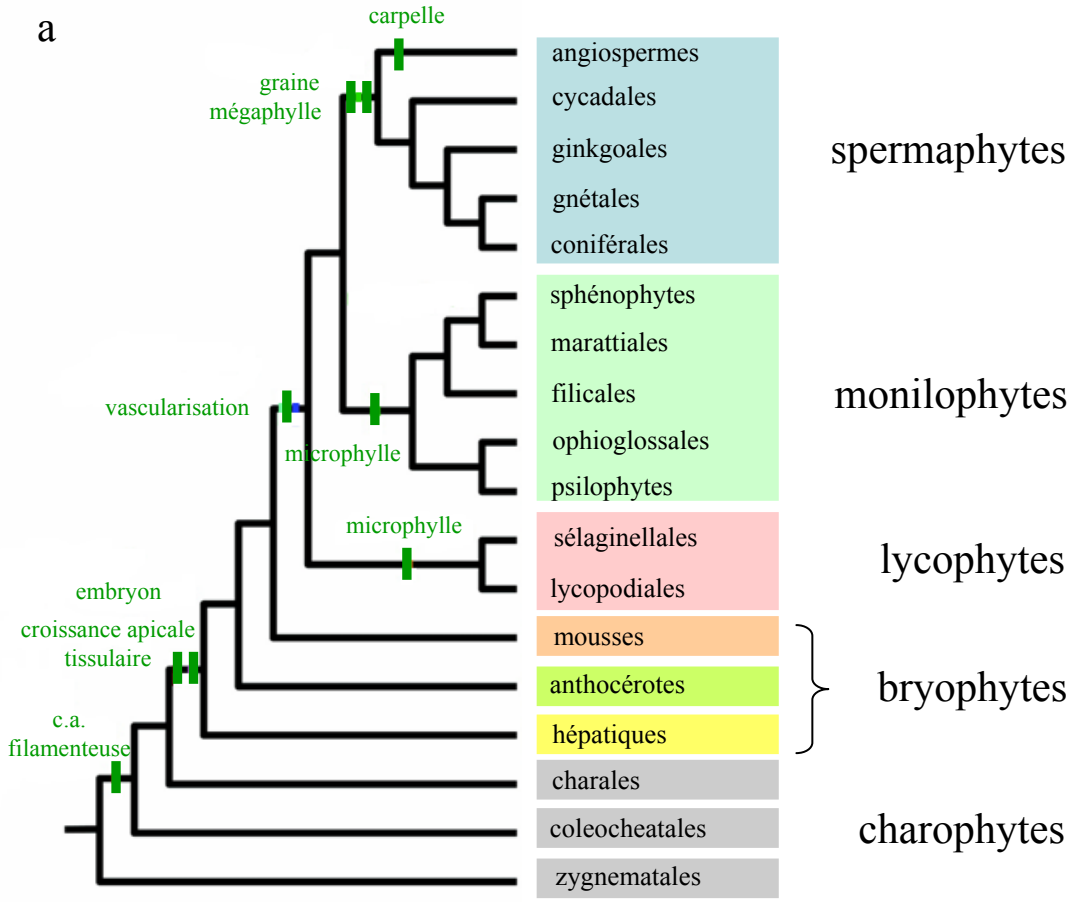
I) Les radiations évolutives majeures au sein des embryophytes

1) Phylogénie des embryophytes

Anciennement regroupées sous le terme de cormophytes (du latin *cormus* : tige) ou archégoniates (du grec *arkhaios* : ancien, *gynê* : femme), les embryophytes comprennent l'ensemble des plantes terrestres. Elles sont apparues entre l'Ordovicien moyen et le Silurien inférieur (480-430 Ma) et dérivent d'un ancêtre de type algue verte (charophyte). Les embryophytes sont définies par la présence d'un embryon végétal, structure qui se développe après la fécondation à partir du zygote et qui contient un pied ou suçoir, un suspenseur et une tête. La tête de l'embryon se développera par la suite en sporophyte avec racine, tige et feuille.

Malgré de nombreux travaux depuis maintenant une dizaine d'années, la relation entre les différents lignages des embryophytes reste controversée. Ces études visant à reconstruire la phylogénie des embryophytes ont fait appel à différentes approches. Les premiers arbres obtenus étaient basés sur des matrices de caractères morphologiques et plaçaient les hépatiques comme première branche de l'arbre (Graham, 1993; Mishler and Churchill, 1984; Mishler and Churchill, 1985; Mishler et al., 1994). Le principal biais de ces reconstructions résidait en l'homoplasie, à savoir la prise en compte de caractères similaires mais non homologues (ne provenant pas d'un ancêtre commun) chez différentes espèces. Par conséquent, les méthodes de reconstruction phylogénétiques ont pris en compte des matrices de caractères moléculaires sans toutefois résoudre la base de l'arbre des embryophytes (Groth-Malonek et al., 2005; Lewis et al., 1997; Nickrent et al., 2000; Nishiyama et al., 2004; Qiu et al., 1998; Turmel et al., 2006; Wolf et al., 2005).

Ces difficultés pour reconstruire la phylogénie des embryophytes sont le résultat de différents phénomènes parmi lesquels on peut citer : une trop grande divergence entre les embryophytes et le groupe externe (les charophytes), un registre fossile lacunaire, un taux d'évolution fortement hétérogène entre les lignages, et tout simplement un nombre insuffisant de sites informatifs ou un échantillonnage taxonomique biaisé (Graybeal, 1998) (Delsuc et al., 2005). Les phylogénies ultérieures ont élargi l'échantillonnage taxonomique et augmenté le nombre de marqueurs moléculaires pris en compte (Leebens-Mack et al., 2005; Korall et al., 2006). La dernière étude phylogénomique en date repose sur l'utilisation de 4 marqueurs chloroplastiques, 2 mitochondriaux et 193 gènes nucléaires (Qiu et al., 2006). Cette phylogénie robuste est considérée comme le consensus actuel dans ce manuscrit **(figure 1a)**.



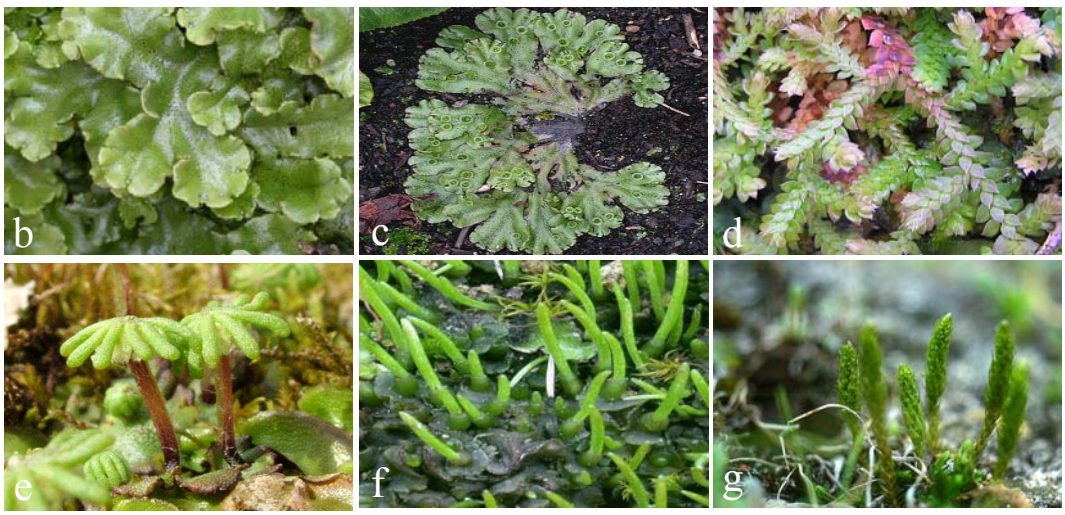


Figure 1 : Phylogénie moléculaire des embryophytes et illustration des premières branches de l'arbre. La phylogénie des embryophytes obtenue par Qiu et *al.* (2006) place respectivement les hépatiques, les anthérocérotes puis les mousses (sens strict) à la base du clade des embryophytes ; viennent ensuite les lycophytes. En vert sont indiquées les principales synapomorphies ou traits caractéristiques des grands groupes d'embryophytes (a). Les photos illustrent les phases gamétophytiques (b,c,d) et sporophytiques (e,f,g) respectives de *Marchantia sp.*, *Anthoceros sp.* et *Selaginella sp.*

2) Synapomorphies des principaux phyla

Dans un souci de clarté taxonomique, les embryophytes seront divisés arbitrairement en quatre phyla qui, par définition, ne constituent pas des groupes naturels mais plutôt des regroupements basés sur des critères biologiques (plan d'organisation, mode de nutrition, etc…).

a) Les bryophytes

Le terme bryophyte *lato sensu* s'applique aux trois phyla de plantes terrestres qui sont apparus précocement au cours de l'évolution : les hépaticophytes (hépatiques) **(figures 1b, 1e)**, les anthocérotophytes (anthocérotes) **(figures 1c, 1f)** et les bryophytes (mousses) *stricto sensu*. Les bryophytes ne possèdent pas de vrai système vasculaire (ni xylème, ni phloème) qui n'apparaît que chez les trachéophytes. Certains cependant ont des tissus conducteurs spécialisés, mais qui ne sont pas lignifiés : les leptoïdes et les hydroïdes. Les pseudophylles des bryophytes sont distinctes des feuilles vraies de par leur histologie. De même, les rhizoïdes des bryophytes ne sont pas homologues aux racines des trachéophytes. De plus, ces organes diffèrent dans le sens où les rhizoïdes ont uniquement une fonction de fixation au substrat alors que les racines interviennent également dans l'absorption de l'eau et des nutriments.

Le cycle de reproduction des bryophytes est diplo-haplophasique avec une forte prédominance de la génération gamétophytique (n) sur la génération sporophytique (2n) **(figure 2a)**. Ainsi, le gamétophyte en forme de thalle (hépatiques, anthocérotes) ou de plante feuillée (mousses) représente la partie la plus visible de la plante. Il différencie des organes sexuels mâles (anthéridies) qui produisent les anthérozoïdes et des organes sexuels femelles (archégones) produisant l'oosphère. La fécondation par zoïdogamie, en nécessitant de l'eau pour la mobilité de l'anthérozoïde, est dépendante du milieu extérieur. De la fusion entre un anthérozoïde et une oosphère naît le sporophyte qui se développe en parasite sur le gamétophyte. Ce sporophyte est très réduit et ne différencie qu'un type de sporange (homosporangie) dans lequel a lieu la méiose et la formation de spores identiques (homosporie). La germination d'une spore conduira à un protonéma qui donnera un nouveau gamétophyte.

b) Les ptéridophytes

Le terme ptéridophyte englobe les quatre embranchements actuels suivants : les lycopodiophytes (lycopodes) **(figures 1d, 1g)**, les équisétophytes (prêles), les psilophytes

(psilotums) et les filicophytes (fougères). D'un point de vue évolutif, les premiers lignages de ptéridophytes marquent l'apparition des plantes vasculaires. Au niveau de la reproduction, les ptéridophytes se dispersent via des spores et non encore par l'intermédiaire de graines comme il est le cas chez les spermaphytes.

Le cycle biologique des ptéridophytes marque l'apparition d'une phase sporophytique dominante **(figure 2b)**. Ainsi, la plante feuillée porte les sporanges abritant les spores, spores qui donneront à leur tour naissance à un gamétophyte différent de celui des bryophytes : le prothalle. Le prothalle est une petite lame mince formée de cellules chlorophylliennes qui porte les anthéridies et/ou archégones.

Chez les filicophytes, les sporanges sont portés par la face inférieure des frondes, groupés en sores et sont tous identiques (homosporangie). Chez les lycopodiophytes, on distingue d'une part les macrosporophylles (plus grandes) situées à la base de l'épi et portant les sporanges de plus grande taille (macrosporanges), d'autre part les microsporophylles (plus petites) situées au sommet portant des sporanges plus petits (microsporanges). Il y a donc hétérosporangie chez ces espèces.

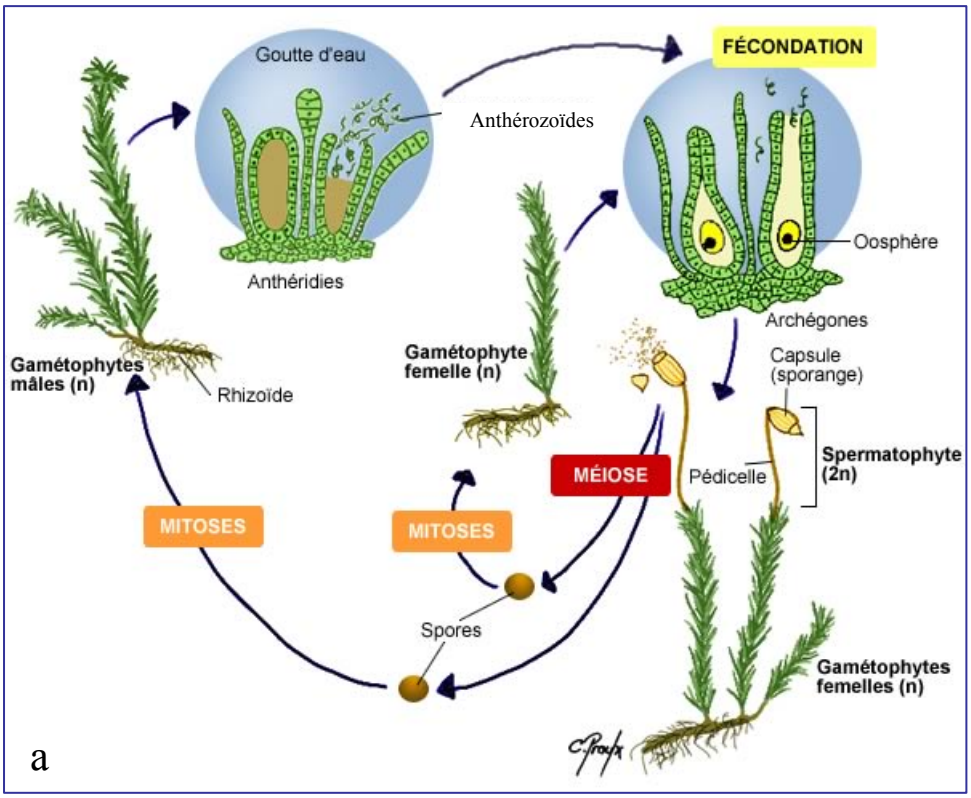
Dans les sporanges, il y a réduction chromatique et production de méiospores qui sont de même taille (homosporie comme chez les filicophytes) ou de tailles différentes (microspores et macrospores comme chez les équisétophytes et lycopodiophytes). Dans le cas des homospores, la germination donne un prothalle bisexué (homoprothallie) alors qu'elle débouche sur deux prothalles unisexués différents (hétéroprothallie) chez les hétérospores. Soulignons ici une innovation évolutive chez *Selaginella* : le microprothalle reste entièrement inclus dans la microspore (endoprothallie totale) et le macroprothalle est partiellement inclus dans la macrospore (endoprothallie partielle). La fécondation nécessite toujours la présence d'eau dans le milieu pour favoriser la rencontre des gamètes.

c) Les gymnospermes

Le terme gymnosperme est utilisé pour parler des cinq embranchements suivants : les cycadophytes (ex. *Cycas*), les ginkgophytes (ex. *Ginkgo*), les conifèrophytes I (ex. *Pinus*), les conifèrophytes II (ex. *Taxus*), ainsi que les gnétophytes (ex. *Gnetum*). Les gymnospermes se caractérisent par le développement du xylème secondaire et sont donc par conséquent toutes ligneuses. Au niveau du xylème, les trachéides ont à la fois un rôle de transport et de soutien.

Le cycle de reproduction des gymnospermes est toujours diplo-haplophasique avec le sporophyte représenté par la plante feuillée et le gamétophyte réduit à quelques cellules

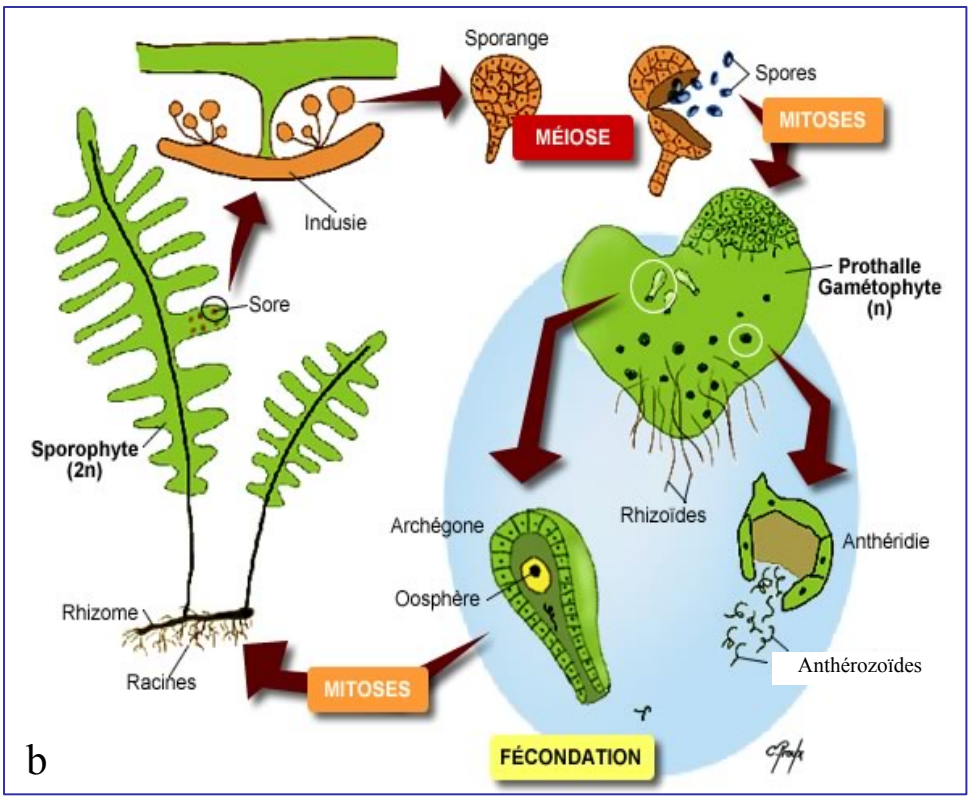
bryophytes (ex : *Polytrichum sp*)



PHASE (2n) DOMINANTE

PROTHALLIE

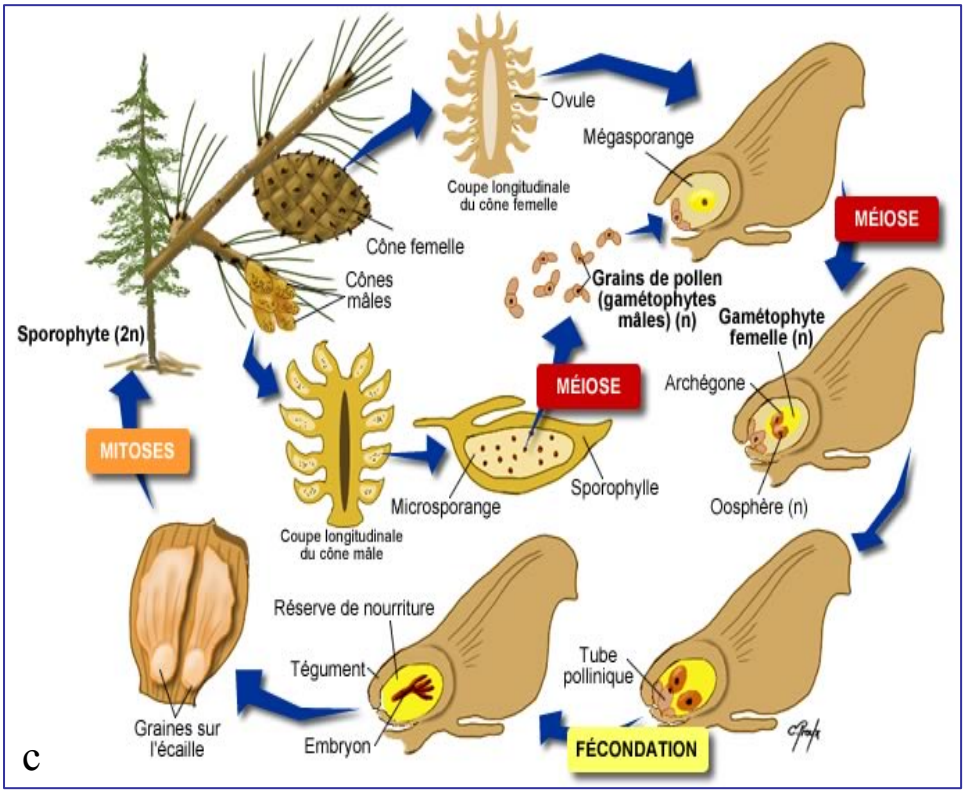
ptéridophytes (ex : *Polypodium sp*)



SIPHONOGAMIE

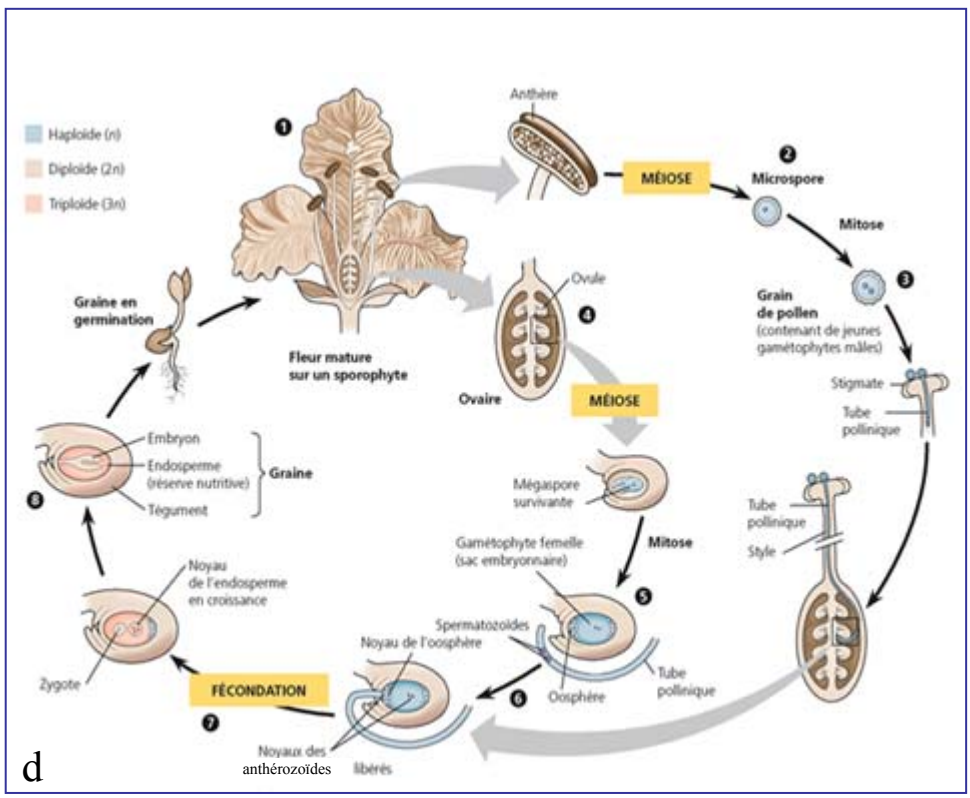
ENDOPROTHALLIE

GRAINE



gymnospermes
(ex : *Pinus sp*)

DOUBLE FÉCONDATION



DISPARITION :

- ARCHÉGONE
- ANTHÉRIDIE

angiospermes
(ex : *Iris sp.*)

Figure 2 : Comparaison des cycles de vie des bryophytes, ptéridophytes, gymnospermes et angiospermes. Les illustrations ont été modifiées à partir de celles réalisées par le professeur Chantal Proulx (*http://www.colvir.net/prof/chantal.proulx/images/reproduction*).

(figure 2c). Les organes reproducteurs, regroupés en cônes mâle et femelle, sont constitués par les étamines et les ovules :

- Étamine : organe mâle correspondant à la microsporophylle avec le microsporange appelé aussi sac pollinique. Elle donne les grains de pollen ou gamétophytes mâles.

- Ovule : organe femelle correspondant au macrosporange (qui prend le nom de nucelle) dans lequel se forme après méiose quatre macrospores. Trois dégénèrent, la macrospore restante se développe au sein du nucelle par mitose pour donner le prothalle femelle ou endosperme. Ce tissu de réserve (n) possède des archégones simplifiés renfermant les oosphères.

Emportés par le vent, les grains de pollen arrivent à la surface du nucelle via le micropyle de l'ovule. Chaque grain de pollen germe et émet un tube pollinique qui pénètre le nucelle jusqu'à une oosphère (fécondation de type siphonogamie). Ce tube pollinique transporte deux noyaux végétatifs et deux anthérozoïdes dont l'un des deux féconde le gamète femelle après fusion avec l'oosphère (gamète femelle). L'ovule fécondé, avec son embryon et ses réserves, forme la graine qui représente une innovation majeure chez les spermaphytes.

d) Les angiospermes

D'un point de vue anatomique, les angiospermes possèdent des vaisseaux spécialisés dans le transport des sèves. Cependant, les angiospermes sont essentiellement caractérisés par des innovations au niveau de leur appareil reproducteur.

Le cycle de vie des angiospermes ressemble globalement à celui des gymnospermes avec néanmoins des différences majeures au niveau des organes femelles et des mécanismes de la fécondation **(figure 2d)**. Par définition de l'angiospermie, le gamétophyte femelle (sac embryonnaire) se trouve totalement internalisé dans le macrosporange (nucelle), lui-même enfermé dans les tissus carpellaires. La structure de type archégone n'est plus reconnaissable chez les angiospermes. Les premières étapes de la germination pollinique sont similaires à celles des gymnospermes. Au contact du sac embryonnaire, le tube pollinique décharge le noyau végétatif dégénère et le noyau reproducteur se divise et donne naissance à deux gamètes (non ciliés) dont l'un féconde l'oosphère (siphonogamie) pour former l'embryon (2n) et l'autre féconde les deux noyaux polaires pour former un tissu de réserve ou albumen (3n). Ce phénomène est décrit sous le terme de double fécondation et représente une synapomorphie des angiospermes. L'albumen se développe au détriment du nucelle. L'embryon continue son développement et forme la jeune plantule (futur sporophyte) au sein de la graine.

3) Innovations évolutives majeures

a) Données morphologiques

La combinaison de la phylogénie et des synapomorphies des principaux embranchements des embryophytes est très informative dans le sens où elle permet de placer les innovations clé sur les branches de l'arbre **(figure 1a)**.

b) Données génomiques

L'année 2008 marque le début de l'ère génomique pour les lignages ayant divergés très tôt au cours de l'évolution des embryophytes. En effet, le génome de *Physcomitrella patens* (bryophyte) et celui de *Selaginella moellendorffii* (lycophyte) ont été séquencés et publiés récemment (Rensing et al., 2008). Ces nouveaux génomes comblent partiellement un manque de données génomiques au sein des embryophytes, pour lesquels la répartition des espèces séquencées était fortement biaisée envers les angiospermes : *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa* et *Vitis vinifera*. Comme je vous l'ai rappelé précédemment, il y a endoprothallie partielle ou totale à la fois chez *Selaginella* (lycophyte) et chez l'ensemble des spermaphytes. S'agit-il d'un phénomène qui s'est produit dans le lignage précédant la spéciation des lycophytes et qui aurait été perdu chez les filicophytes ? Ou avons-nous à faire à des phénomènes de convergence évolutive à la fois chez Selaginella et les spermaphytes ?

II) L'apparition des angiospermes ou l' "abominable mystère" de Darwin

La formation très rapide, à l'échelle des temps géologiques, d'une biodiversité d'une richesse inouïe du point de vue des formes et des modes de vie a longtemps constitué pour les biologistes un "abominable mystère", selon la formule utilisée par Charles Darwin pour décrire son incapacité à comprendre les causes du phénomène.

1) Données fossiles

a) Gymnospermes candidats comme ancêtres des angiospermes

On peut distinguer plusieurs lignages éteints de gymnospermes présentant des caractères anatomiques précurseurs de ceux des angiospermes **(figure 3)** :

- les Glossoptéridales, apparues au carbonifère/permien, ont des organes reproducteurs qui sont portés par la face ventrale de feuilles modifiées. Le genre *Denkania* est particulièrement intéressant dans le sens où il possède des ovules bitegminés comme chez les

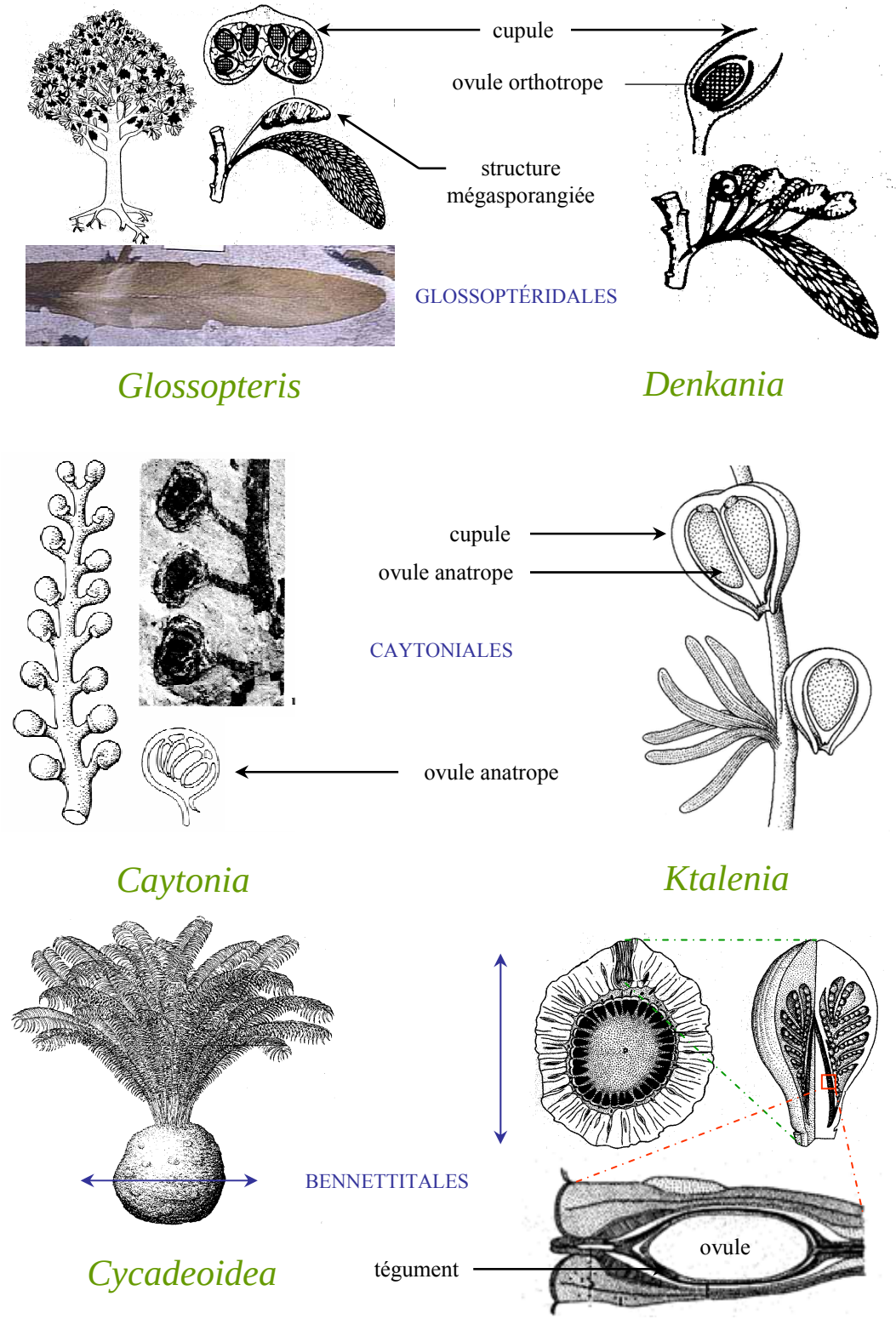


Figure 3 : Fossiles de gymnospermes ptéridospermes. Sont illustrés l'appareil végétatif et les structures reproductrices femelles des glossoptéridales (*Glossopteris* et *Denkania*, Permien-Trias), des caytoniales (*Caytonia*, Jurassique moyen; *Ktalenia*, Crétacé inférieur) et des bennettitales (*Cycadeoidea*, Jurassique inférieur). D'après Leroy (1993).

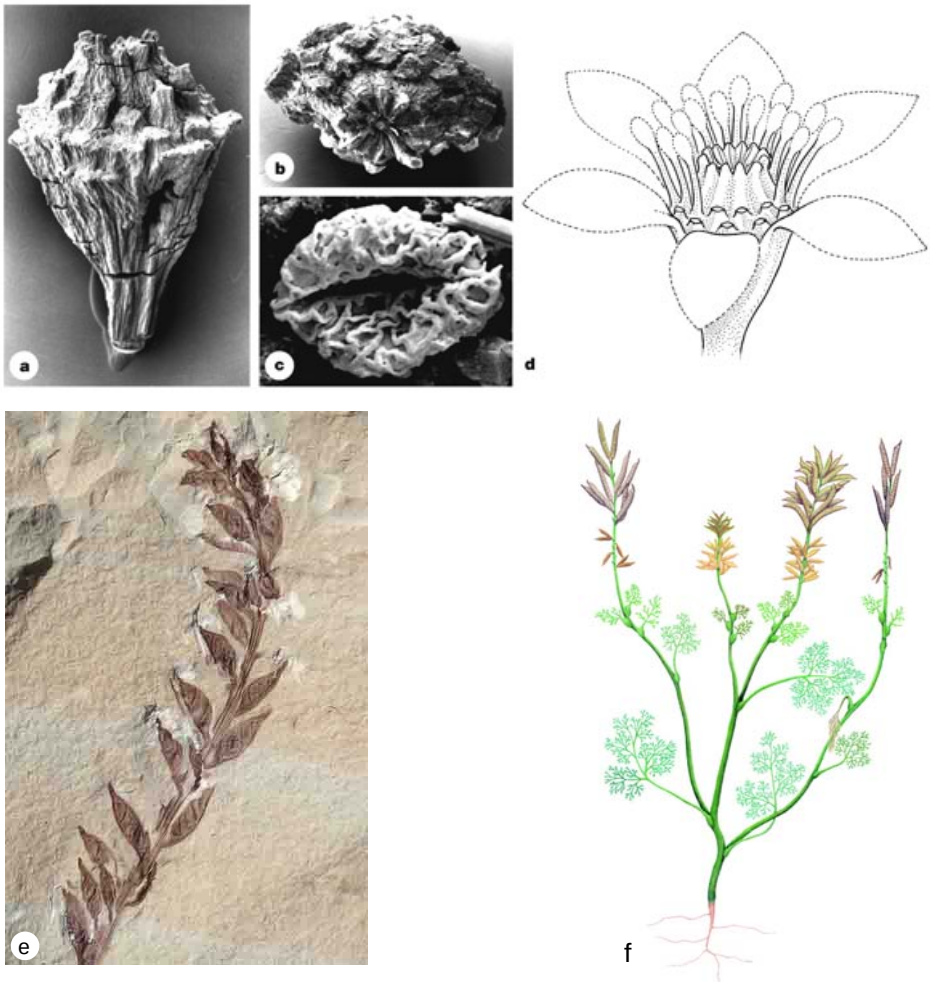


Figure 4 : Fossiles et reconstitutions schématiques des premières angiospermes. La plus ancienne angiosperme connue à ce jour est représentée par un fossile de Nymphaeales découvert au Portugal : (a) fleur fossile, (b) carpelles, (c) grain de pollen, (d) reconstitution de la fleur (Friis et al., 2001). Malgré sa position phylogénétique controversée, le fossile *Archaefructus,* découvert en Chine, pourrait être un des premiers lignages de plantes à fleurs : (e) fossile d'*Archaefructus*, (f) reconstitution du port général (Sun et al., 1998).

angiospermes (Surange and Chandra, 1971). Cependant, ces ovules sont encore orthotropes tels les ovules de gymnospermes.

- les Caytoniales, apparues au jurassique/crétacé inférieur, possèdent des cupules multi ou uniovulées. Encore une fois, il est possible de faire le parallèle entre la cupule et le tégument externe des ovules chez les plantes à fleurs. Parmi les Caytoniales, la famille des Corystospermaceae est particulièrement intéressante sur le plan évolutif car ses genres présentent des cupules recourbées vers la tige, mimant ainsi l'ovule anatrope des angiospermes. Les singularités anatomiques de cette famille servent d'arguments à la théorie "Mostly Male" qui s'attelle à l'origine des plantes à fleurs (voir § II.3.a).

- les Bennettitales, apparues au jurassique inférieur, présentent des similitudes avec les Gnétales au niveau de l'organisation bisexuée et la position axillaire des appareils reproducteurs bisporangiés, de l'appareil végétatif à ramification dichasiale, et des stomates (voir théorie antophyte § II.2.a).

b) Premières angiospermes

Un autre pan de la paléobotanique s'intéresse non pas aux protoangiospermes mais aux fossiles des premières plantes à fleurs. Ces dernières années, la découverte du nouveau genre fossile *Archaefructus* a fait couler beaucoup d'encre **(figures 4e-f)**. Il s'agit de fossiles découverts en Chine dont l'excellent état de conservation a permis la description de racines, de tiges, de feuilles, de fleurs et de fruits (Sun et al., 1998). *Archaefructus* serait une petite plante aquatique possédant des fleurs dépourvues de pièces périanthaires et de bractées, dont les étamines et les carpelles sont répartis le long de l'axe reproducteur. Le genre *Archaefructus* a été suggéré comme étant un précurseur des plantes à fleurs actuelles ayant été trouvé dans une couche géologique datée du jurassique supérieur (Sun et al., 1998) (Sun et al., 2002). Selon cette hypothèse, les premières angiospermes seraient apparues avant le Crétacé. Cependant, il semblerait que le gisement Yixian date plutôt du crétacé supérieur ce qui déloge *Archaefructus* de la place de fleur ancestrale (Friis et al., 2003). Le réexamen morphologique de ce fossile plaide plutôt pour une plante dont le mode de vie aquatique aurait conduit à la perte secondaire du périanthe et des bractées. Les fleurs petites et unisexuées seraient regroupées en une inflorescence (Friis et al., 2003) (Friis et al., 2005). La position phylogénétique du genre *Archaefructus* reste actuellement incertaine.

La découverte récente du fossile *Schmeissneria* apporte de nouveaux questionnements. Découvert en Chine et daté du Jurassique moyen, il possède des caractères angiospermiens comme la présence d'un carpelle clos (Wang et al., 2007). Cependant, les auteurs restent

prudents et évoquent soit une origine pré-Crétacé des angiospermes, soit un nouveau lignage de plantes à graines dont *Schmeissneria* serait l'unique représentant connu à ce jour.

Finalement, la plus ancienne plante à fleur connue à ce jour (et sans ambiguïté) est représentée par un fossile de Nymphaeales **(figures 4a-d)**. Un seul spécimen a été découvert au Portugal dans des gisements datés du Crétacé inférieur (Friis et al., 2001). La fleur est minuscule, à symétrie radiale, et vraisemblablement hermaphrodite. Tous les organes floraux ont disparu à l'exception des carpelles, laissant des marques interprétées comme les sites d'insertion des étamines. Le gynécée est composé de 12 carpelles soudés et arrangés en un verticille qui entoure une dépression apicale. Ce renfoncement est retrouvé uniquement chez certaines espèces de Nymphaeaceae et d'Illiciaceae actuelles (Endress and Igersheim, 2000). La présence de Nymphaeales dans le gisement Vale de Aqua est également supportée par des empreintes foliaires (Saporta, 1894), ainsi que des téguments externes de graines (Friis et al., 1999) (Friis et al., 2000).

2) Phylogénie des angiospermes actuelles

a) Fin de la théorie anthophyte

Les gnétales présentent des similarités morphologiques assez remarquables avec les angiospermes comme la présence de vaisseaux dans le xylème, de feuilles à symétrie bilatérale, de structures reproductrices ressemblant à des fleurs, ainsi qu'une double fécondation (Friedman, 1990). En parallèle, la possibilité d'un lien de parenté entre les bennettitales et les gnétales a été avancée sur des caractères morphologiques. L'ensemble de ces données a conduit à regrouper les gnétales, les angiospermes et les bennettitales en un même clade : les anthophytes (Crane, 1985) (Doyle and Donoghue, 1986).

La reconstruction phylogénétique des spermaphytes et la remise en question de la théorie anthophyte n'a pas été un processus linéaire (Donoghue and Doyle, 2000). Les premières phylogénies moléculaires robustes obtenues ne soutiennent pas l'existence du clade anthophyte (Chaw et al., 2000). Au contraire, elles indiquent que les gymnospermes actuelles sont monophylétiques et placent les gnétales au sein des conifères directement avec les pinaceae, menant à l'hypothèse gnépine (Chaw et al., 2000).

Par conséquent, les caractères morphologiques qui unissaient les gnétales et les angiospermes se révèlent être des convergences évolutives et non des homologies. Ces caractères anatomiques ont été minutieusement réexaminés chez les gnétales. Les vaisseaux du bois des angiospermes dériveraient de trachéides scalariformes alors que les vaisseaux des gnétales ressemblent plus à des trachéides aréolées, comme chez les conifères. Les gnétales

ressemblent également aux conifères dans le sens où le xylème primaire est dénué de perforations scalariformes. Sur le plan reproductif, l'ovule des gnétales est contenu dans une enveloppe non close et qui ménage un micropyle très allongé. Enfin, le second événement de fécondation conduit à la formation d'une structure diploïde et non d'un tissu de réserve triploïde, l'albumen, comme chez les plantes à fleurs (Friedman and Floyd, 2001).

b) Le grade ANA à la base de l'arbre des angiospermes

L'acronyme ANA (anciennement ANITA) résulte du regroupement des ordres suivants : **A**mborellales **(figures 6e-f)**, **N**ymphaeales **(figures 6a-b)** et **A**ustrobaileyales **(figure 6c)**. Il s'agit d'un grade et non d'un clade car les premières branches des plantes à fleurs sont regroupées sur des critères morphologiques et non phylogénétiques **(figure 5)**. Autrement dit, le grade ANA n'est pas monophylétique. Les dernières phylogénies moléculaires sont congruentes pour placer le grade ANA à la base des angiospermes. La recherche de signatures moléculaires ou d'événements moléculaires rares (indels, duplications, insertions d'éléments transposables, etc…) se révèle être un bon outil phylogénétique. Par exemple, la présence d'un indel dans le gène *matR* (Qiu et al., 1999) et dans le gène *APETALA3* (Vandenbussche et al., 2003) a permis de différencier les espèces du grade ANA du reste des angiospermes.

La position basale du grade ANA a cependant été remise en question par des travaux ultérieurs (Goremykin et al., 2003; Goremykin et al., 2004). Ces travaux reposent sur la concaténation de 61 gènes présents dans 14 génomes chloroplastiques séquencés chez les plantes terrestres (dont ceux du grade ANA : *Amborella trichopoda* et *Nymphaea alba*). Dans la phylogénie ainsi obtenue, les monocots sont plus basales que le grade ANA. Cependant, les arbres obtenus semblent biaisés par un échantillonnage restreint à la famille des Poaceae chez les monocots (Soltis and Soltis, 2004; Stefanovic et al., 2004).

Des données récentes montrent que la phylogénie actuelle des angiospermes est en train d'évoluer, notamment au niveau du grade ANA. Tout d'abord, la position relative des Amborellales et des Nymphaeales reste controversée et continue à alimenter le débat (Barkman et al., 2000; Graham and Olmstead, 2000; Moore et al., 2007; Zanis et al., 2002). De manière intéressante, une étude récente vient de montrer que les Hydatellaceae se branchent parmi les premiers lignages de l'arbre phylogénétique des angiospermes, et plus particulièrement avec les Nymphaeales (Saarela et al., 2007). Il s'agit d'une famille réduite (2 genres : *Hydatella* et *Trithuria*) de plantes aquatiques minuscules qui présentent des fleurs

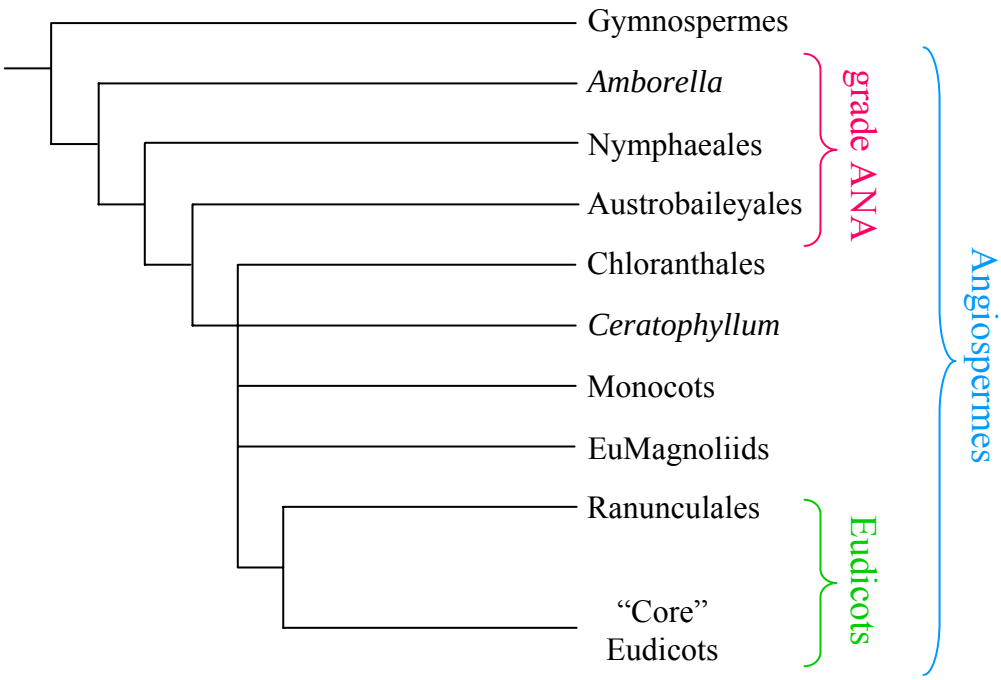


Figure 5 : Arbre phylogénétique simplifié des plantes à fleurs. Les Gymnospermes sont utilisés ici comme groupe externe pour enraciner l'arbre. Notons la position très basale des espèces du grade ANA composé des groupes ayant divergé le plus tôt au cours de l'évolution : ***A**mborella*, **N**ymphaeales, Illiciales et Schisandraceae, Trimeniaceae et **A**ustrobaileyaceae (Qiu et al., 1999).

Figure 6 : Illustrations de certaines espèces du grade ANA. (a) Appareil végétatif de *Cabomba aquatica*, (b) bourgeon floral émergé de *C. aquatica*, (c) fleur d'*Illicium anisatum*, (d) plante entière de *Trithuria submersa*, (e) fleur mâle d'*Amborella trichopoda*, (f) fleurs femelles d'*A. trichopoda*.

simples **(figure 6d)**. Ces genres étaient auparavant groupés au sein des Poales, un des groupes les plus dérivés des monocots.

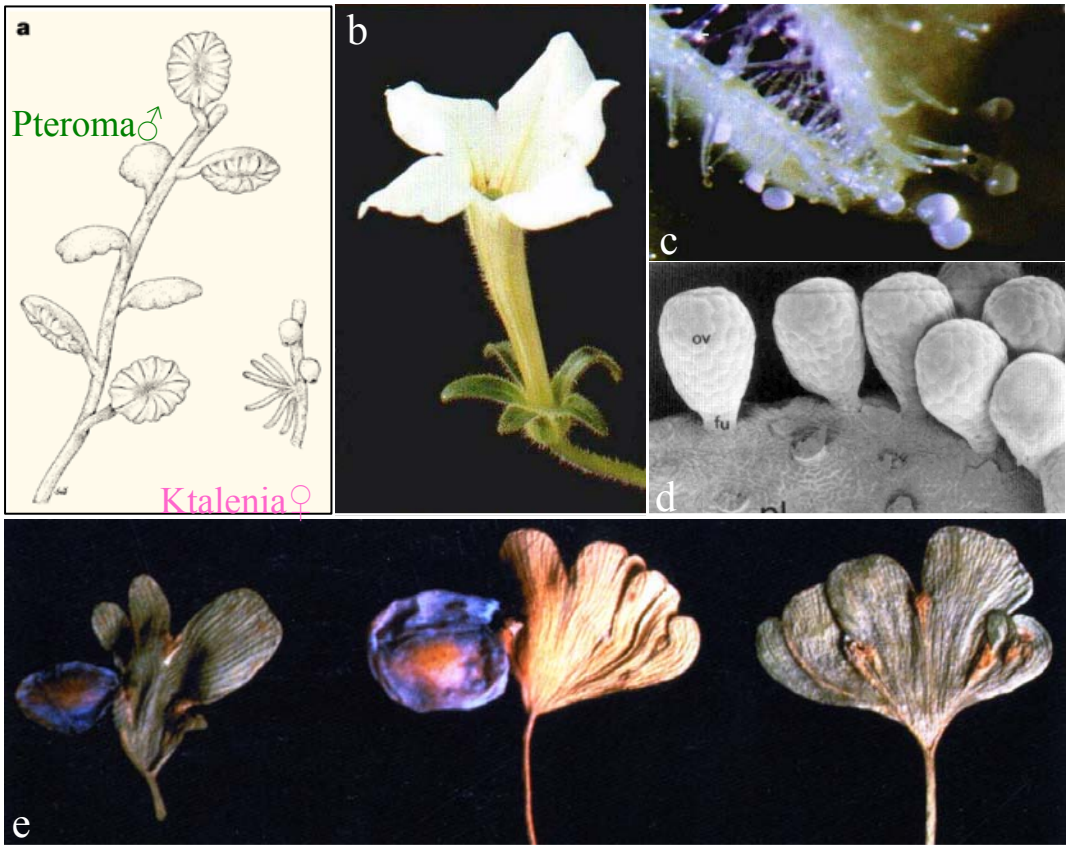
3) Théories actuelles

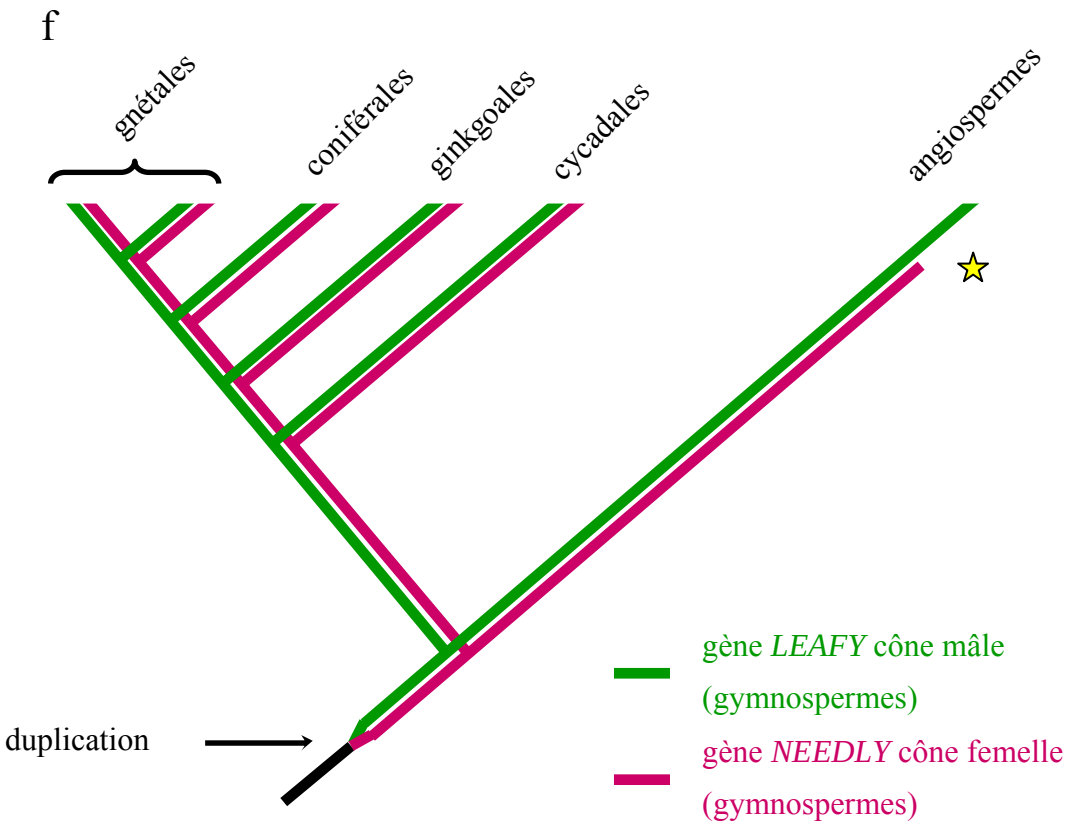
L'apparition de la fleur a nécessité deux grandes étapes évolutives dont la chronologie relative reste indéterminée. D'une part, le rapprochement des sexes sur un même axe pour passer d'une structure unisexuée vers un organe bisexué. D'autre part, la fermeture de la feuille carpellaire préexistante autour des ovules pour donner le carpelle. Les théories actuelles sur l'origine des plantes à fleurs s'attachent à l'une ou l'autre de ces étapes.

a) Théorie "Mostly Male"

Cette théorie cherche à expliquer toutes les étapes menant à l'apparition de la fleur même si elle se concentre principalement sur le développement d'organes reproducteurs femelles sur un cône ancestral mâle. La théorie "Mostly Male" considère que la fleur résulte de l'apparition d'ovules ectopiques sur une structure ancestrale de type microsporophylle qui se serait refermée pour donner le carpelle (Frohlich, 2003; Frohlich and Parker, 2000).

Cette théorie repose sur la synthèse d'arguments à la fois moléculaires, paléontologiques, et botaniques. Chez *A. thaliana*, le gène *LEAFY* (*LFY*) est connu pour déclencher l'induction florale par (i) intégration des stimuli externes (Blazquez and Weigel, 2000), (ii) répression des gènes d'identité méristématique *TERMINAL FLOWER1* (*TFL1*) et *AGAMOUS-LIKE24* (*AGL24*) (Liljegren et al., 1999; Parcy et al., 2002; Yu et al., 2004), et (iii) activation des gènes d'identité florale (Bowman et al., 1993; Busch et al., 1999; Kempin et al., 1995; Lohmann et al., 2001; Parcy et al., 2002; Ratcliffe et al., 1999; Wagner et al., 1999; Yu et al., 2004). La reconstruction phylogénétique du gène *LFY* montre qu'il est unique chez les angiospermes alors que deux paralogues, *LFY* et *NEEDLY*, sont présents chez les gymnospermes **(figure 7f)** (Frohlich and Meyerowitz, 1997). Chez les gymnospermes, *LFY* est exprimé dans les cônes mâles et les méristèmes végétatifs (Mellerowicz et al., 1998) alors que *NEEDLY* est exprimé uniquement dans les cônes femelles (Mouradov et al., 1998). Le fait que *LFY* code le programme mâle chez les gymnospermes, et qu'il ait été conservé pour établir les programmes mâle et femelle des fleurs hermaphrodites, suggèrent que la fleur dérive d'une structure mâle. Le développement d'ovules ectopiques est un phénomène bien connu des généticiens moléculaires de la fleur. Par exemple, la surexpression du gène *FPB11* entraîne la formation d'ovules ectopiques sur les sépales et pétales chez *Petunia hybrida* **(figures 7b-d)** (Colombo et al., 1995). De manière moins artificielle, on connaît des individus


a
Pteroma♂
Ktalenia♀
b
c
ov
fu
d
e


f
gnétales
coniférales
ginkgoales
cycadales
angiospermes
duplication
gène *LEAFY* cône mâle (gymnospermes)
gène *NEEDLY* cône femelle (gymnospermes)

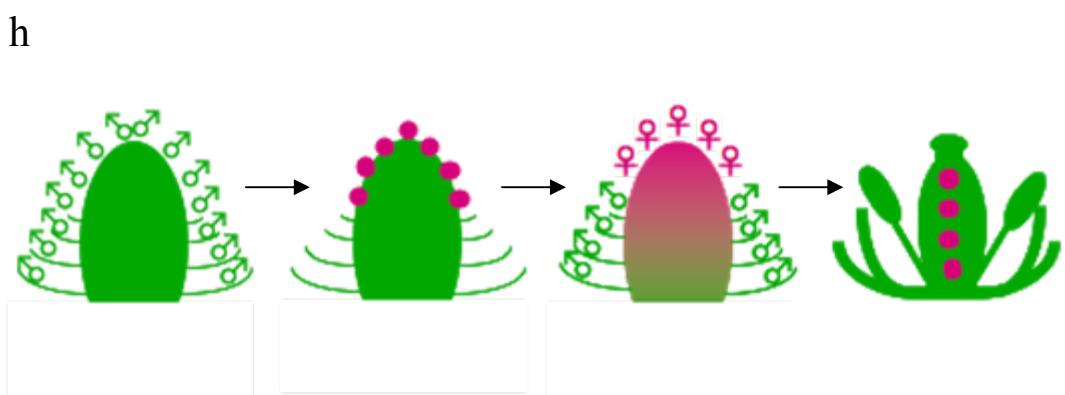


Figure 7 : Données biologiques à partir desquelles fut élaborée la théorie du "Mostly Male''. Selon ce scénario évolutif, la fleur dériverait d'une structure ancestrale mâle qui aurait évolué par les changements successifs suivants : (1) apparition d'ovules (ou cupules) ectopiques sur les microsporophylles, (2) perte des structures mâles sur les microsporophylles avec ovules, (3) fermeture des feuilles modifiées pour donner la feuille carpellaire (g,h). Cette théorie repose d'abord sur des arguments paléontologiques selon lesquels les corytospermales seraient des ancêtres potentiels des angiospermes. On a été retrouvé des fossiles d'individus mâles du genre *Pteroma* et femelles du genre *Ktalenia* présentant des caractères intéressants comme des microsporophylles peltées et des cupules recourbées respectivement (a). En parallèle, des cas d'apparition d'ovules ectopiques ont été reportés, qu'ils soient d'origine artificielle (b, c, d : mutation du gène FBP11 chez *Petunia hybrida*) ou naturelle (e : *Ginkgo biloba*, Cincinnati, USA). Enfin, des données moléculaires sur l'évolution du gène *LFY* chez les spermaphytes renforcent l'idée d'une structure initiale de type mâle.

de *G. biloba* présentant des ovules ectopiques sur leurs feuilles végétatives **(figure 7e)**. Ces exemples suggèrent que l'apparition d'ovules ectopiques n'est pas un phénomène impossible et qu'il peut donc être pris en compte dans une théorie évolutive. Sur un plan paléontologique, les fossiles de corystospermales présentent des microsporophylles simples, non lobées et arrangées en spirale sur la tige (Yao et al., 1995), ce qui correspond exactement à la morphologie requise par la théorie pour rendre compte de l'ancêtre commun des angiospermes **(figures 7a, 7g)**. De plus, les corystospermales montrent une organisation de l'ovule souvent considérée comme les prémices de l'organisation de l'ovule des angiospermes. En effet, les ovules des plantes à fleurs ont deux téguments, alors que les ovules des gymnospermes n'en possèdent qu'un seul. De plus, l'ovule est anatrope (le micropyle est tourné vers le funicule) chez la grande partie des angiospermes, alors qu'il est toujours orthotrope (le micropyle aligné avec la chalaze et le hile) chez les gymnospermes. Chez les corystospermales, un ou deux ovules similaires à ceux des gymnospermes sont enfermés dans une feuille modifiée - la cupule – qui est courbée vers la base de cette cupule (Klavins et al., 2002). Il est suggéré que la cupule des corystospermales pourrait être homologue au tégument externe de l'ovule des angiospermes (Stebbins, 1974).

b) Théorie "Out of Male/Out of Female"

Cette théorie cherche à expliquer la transition de structures unisexuées (cônes) vers une structure hermaphrodite (fleur). La théorie "Out of Male/Out of Female" stipule que l'expression différentielle des gènes d'identité florale de classe B pourrait constituer le mécanisme de détermination du sexe de toutes les spermaphytes (Theissen et al., 2000; Winter et al., 1999). Ainsi, la transition d'une identité mâle vers une identité femelle dépendrait du changement de l'activité d'un nombre restreint de gènes (voire d'un seul gène). Intuitivement, deux possibilités se présentent à nous : le cône ancestral était soit mâle, soit femelle. L'hypothèse "Out of Male" part du principe que les fleurs hermaphrodites avec leur(s) verticille(s) d'organes mâles en périphérie (en bas du cône) et leur verticille d'organes femelles au centre (en haut du cône) proviennent d'un cône mâle. La réduction de l'expression des gènes de classe B (ou l'expression ectopique des gènes B_{sister}) en haut du cône conduit au développement de structures femelles. Au contraire, l'hypothèse "Out of Female" repose sur une expression ectopique des gènes de classe B (ou une réduction de l'expression des gènes B_{sister}) en bas d'un cône ancestral femelle, résultant en l'apparition de structures mâles.

Cette théorie repose sur la découverte des gènes B_{sister} parmi l'ensemble des spermaphytes. L'histoire a commencé par la caractérisation de gènes à boîte MADS chez la gymnosperme *Gnetum gnemon* (Becker et al., 2000; Winter et al., 1999). Le gène *GGM13* définit un clade B_{sister} paralogue au clade B qui est constitué des gènes B des angiospermes ainsi que du gène *GGM12* de *G. gnemon*. De plus, la protéine prédite GGM13 possède des signatures moléculaires propres aux gènes de classe B et absentes des autres protéines à boîte MADS (Becker et al., 2002). Le clonage du gène *ZMM17* chez *Zea mays* (Becker et al., 2002), groupant au sein du clade B_{sister}, a permis de conclure que l'ancêtre commun des spermaphytes possédait déjà un gène B_{sister} il y a 300 Ma. L'étude de l'expression des gènes B_{sister} a également été un facteur déterminant dans l'élaboration de la théorie "Out of Male/Out of Female". Contrairement aux gènes du clade B exprimés dans les structures reproductrices mâles (ainsi que dans les pétales des angiospermes), les gènes B_{sister} sont majoritairement exprimés dans les organes reproducteurs femelles, spécifiquement dans les ovules et les tissus proches de l'ovule (Becker et al., 2002; Kramer and Irish, 2000).

Enfin, il est intéressant de noter que des cônes hermaphrodites ont été décrits dans des populations naturelles de conifères (Günter Theissen, communication personnelle).

c) Théories alternatives

Les deux théories précédentes ont servi de base de réflexion à l'élaboration de plusieurs modèles. Ainsi, la théorie "Mostly Male" a permis à Albert *et al.* (Albert et al., 2002) de construire un modèle d'apparition de la bisexualité sur l'évolution de la pléiotropie des gènes *LFY* et *NEEDLY*. Chez les gymnospermes, ces deux gènes jouent un rôle dans des modules de développement qui sont génétiquement dépendants. En effet, la fonction C est requise pour le programme femelle alors que les fonctions B et C interviennent dans le développement du programme mâle. Une des hypothèses de ce nouveau modèle est de considérer que l'interaction génétique (i.e. redondance partielle) entre *LFY* et *NEEDLY*, qui stabilise la rétention de ces paralogues chez les gymnospermes actuelles, aurait été moins forte dans le lignage des angiospermes et aurait par conséquent conduit à la perte du gène *NEEDLY*. Contrairement à la théorie "Mostly Male", ce modèle ne débouche pas sur des hypothèses dont on pourrait tester expérimentalement la recevabilité.

De même, la théorie "Out of Male/Out of Female" a été reprise et prolongée par David Baum et Lena Hileman (Baum and Hileman, 2006). Dans l'hypothèse selon laquelle la présence *vs.* l'absence de l'expression des gènes B conduirait à l'identité microsporophylle *vs.* macrosporophylle, une des questions serait de savoir comment la fonction B est réprimée dans

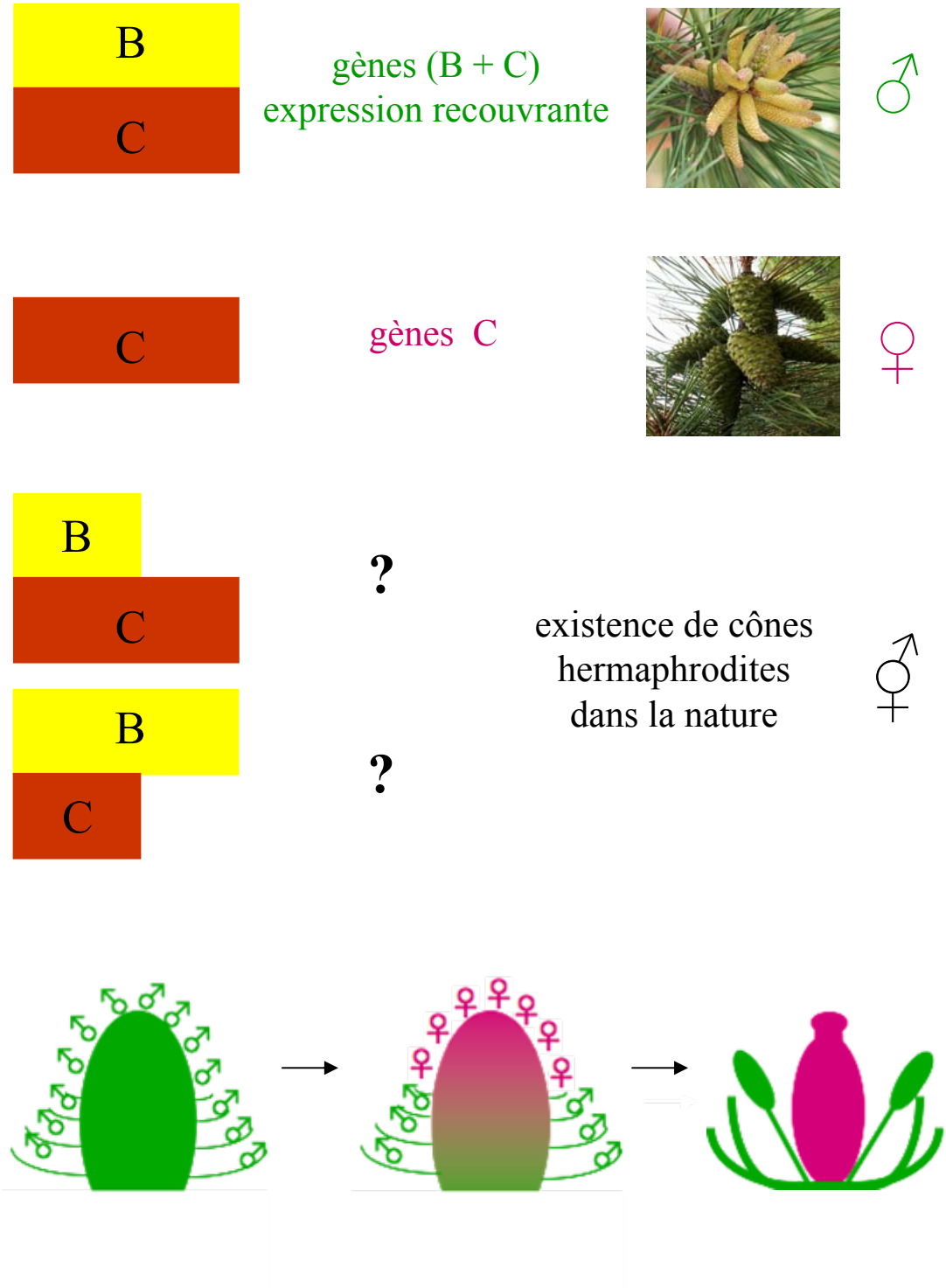


Figure 8 : Détermination de l'identité sexuelle chez les gymnospermes et théorie "Out of Male/Out of Female". Le modèle ABC(D)E est partiellement applicable chez les gymnospermes. Les gènes de classe C sont exprimés uniquement dans les cônes femelles alors que les cônes mâles se caractérisent par une co-expression des gènes de classe B et C. La théorie "Out of Male/Out of Female" stipule que le rapprochement des sexes sur une même structure résulte de modifications dans le patron d'expression des gènes de classe B.

certaines régions du cône ? Il a été suggéré que cette régulation pourrait se faire via la diffusion apico-basale d'une hormone ou autre facteur diffusible (Theissen and Becker, 2004). David Baum et Lena Hileman (Baum and Hileman, 2006) privilégient un modèle basé sur la cinétique de l'expression et la capacité à former des complexes multimériques des gènes B et C pour rendre compte de la répression de la fonction B. Dans un premier temps, une augmentation du niveau protéique de LFY conduirait à une augmentation de la quantité des protéines de classe B et C. À ce moment, les complexes protéiques seraient composés de protéines des classes B, C et E promouvant le développement des microsporophylles. Puis, le modèle fait l'hypothèse que l'expression des gènes B atteindrait son maximum avant celui des gènes C. Ainsi, les complexes protéiques seraient progressivement dépourvus de protéines de classe B (mais toujours de classe C), conduisant au développement des macrosporophylles.

III) La génétique évolutive du développement du carpelle : état de l'art

1) Définition et terminologie du carpelle

Le carpelle est l'unité reproductrice femelle des plantes à fleurs et constitue le quatrième verticille de la fleur **(figure 9a, 10)**. Généralement, on distingue trois parties sur un carpelle (de haut en bas) :

- le stigmate, partie terminale du carpelle recouvert de papilles stigmatiques visqueuses jouant un rôle dans la réception et la reconnaissance des grains de pollen

- le style, partie allongée du carpelle reliant l'ovaire et le stigmate

- l'ovaire, partie basale du carpelle renfermant les ovules qui se transforme en fruit après la fécondation

Le nombre de carpelles est très variable selon les espèces (de un à plusieurs centaines). On parle de syncarpie lorsque les différents carpelles d'une fleur sont soudés entre eux pour donner un pistil, en opposition à l'apocarpie caractérisée par des carpelles libres. Chez *A. thaliana*, le gynécée est composé de deux carpelles fusionnés **(figure 9)**.

2) Identité du domaine carpellaire

Selon le modèle ABC(D)E, l'identité "carpelle" est la résultante de la co-expression de gènes de classe C et E **(figure 10)**. La fonction C consiste en : (i) la détermination de l'identité "étamine" et "carpelle", (ii) la répression de la fonction A et (iii) la croissance déterminée du méristème floral. Seul le gène *AGAMOUS* (*AG*) est connu pour assurer cette triple fonction chez *A. thaliana*. Le mutant fort *ag-1* présente des défauts d'identité des organes floraux dans les troisième et quatrième verticilles où des pétales et des sépales se

développent à la place des étamines et des carpelles (Bowman et al., 1989). Ceci peut s'expliquer par l'absence de répression des gènes de classe A dans les verticilles les plus internes **(figure 10d)** (Gustafson-Brown et al., 1994). De plus, le méristème floral de ces mutants est indéterminé, résultant en la réitération de nombreux sépales et pétales au centre de la fleur (Bowman et al., 1989). La fonction E est assurée par les gènes *SEPALLATA* (*SEP*) au nombre de quatre chez *A. thaliana* (Honma and Goto, 2001). Le quadruple mutant *sep1 sep2 sep3 sep4* est dépourvu de carpelles et présente des organes ressemblant à des feuilles dans le quatrième verticille **(figure 10e)** (Ditta et al., 2004; Pelaz et al., 2000). Malgré le rôle central d'*AG* pour l'identité carpellaire, les tissus du carpelle peuvent tout de même se développer dans le premier verticille de mutants *ap2 ag* (Bowman et al., 1991). Ces travaux suggèrent l'existence d'une deuxième voie de signalisation, indépendante d'*AG*, dans l'acquisition de l'identité "carpelle". Les gènes *SHATTERPROOF1* (*SHP1*) et *SHP2*, paralogues d'*AG*, sont des candidats potentiels dans la mesure où les tissus carpellaires ectopiques disparaissent dans le quadruple mutant *ap2 ag shp1 shp2* (Pinyopich et al., 2003). De plus, des études de complémentation montrent que les protéines SHP et AG sont largement équivalentes d'un point de vue fonctionnel, leurs rôles respectifs résulteraient de différences dans leurs patrons d'expression (Pinyopich et al., 2003).

L'identification de nouveaux gènes impliqués dans la voie *AG* a été rendue difficile à cause d'une forte redondance fonctionnelle et létalité. Cependant, la voie de signalisation *AG* commence à révéler ses secrets. L'accumulation d'ARNm *AG* au centre du méristème floral est essentiellement sous contrôle transcriptionnel. Alors que des régulateurs positifs comme *LEAFY* et *WUSCHEL* activent l'expression du gène *AG* (Busch et al., 1999; Lenhard et al., 2001; Lohmann et al., 2001), des régulateurs négatifs tels qu'*APETALA2*, *LEUNIG*, *SEUSS*, *AINTEGUMENTA* et *STERILE APETALA* veillent à ce qu'*AG* ne soit pas exprimé dans les premier et deuxième verticilles (Byzova et al., 1999; Drews et al., 1991; Krizek et al., 2000; Liu and Meyerowitz, 1995). En parallèle, l'expression du gène *AG* semble être contrôlé au niveau post-transcriptionnel. Les gènes *HUA1*, *HUA2*, *HUA ENHANCER2* (*HEN2*) et *HEN4* jouent un rôle dans le bon déroulement de la maturation du pré-ARNm *AG*, ceci d'une manière spécifique à *AG* et non à l'ensemble des ARNm de la cellule (Cheng et al., 2003). De même qu'*AG*, le gène *HEN1* réprime la fonction A dans les troisième et quatrième verticilles en permettant l'accumulation de microARNs, connus pour réguler négativement le gène AP2 (Chen et al., 2002). Enfin, les gènes *PAUSED* et *HEN3* interviennent dans la détermination des étamines et des carpelles, mais leurs fonctions dans la voie *AG* ne sont pas encore élucidées (Li and Chen, 2003; Wang and Chen, 2004).

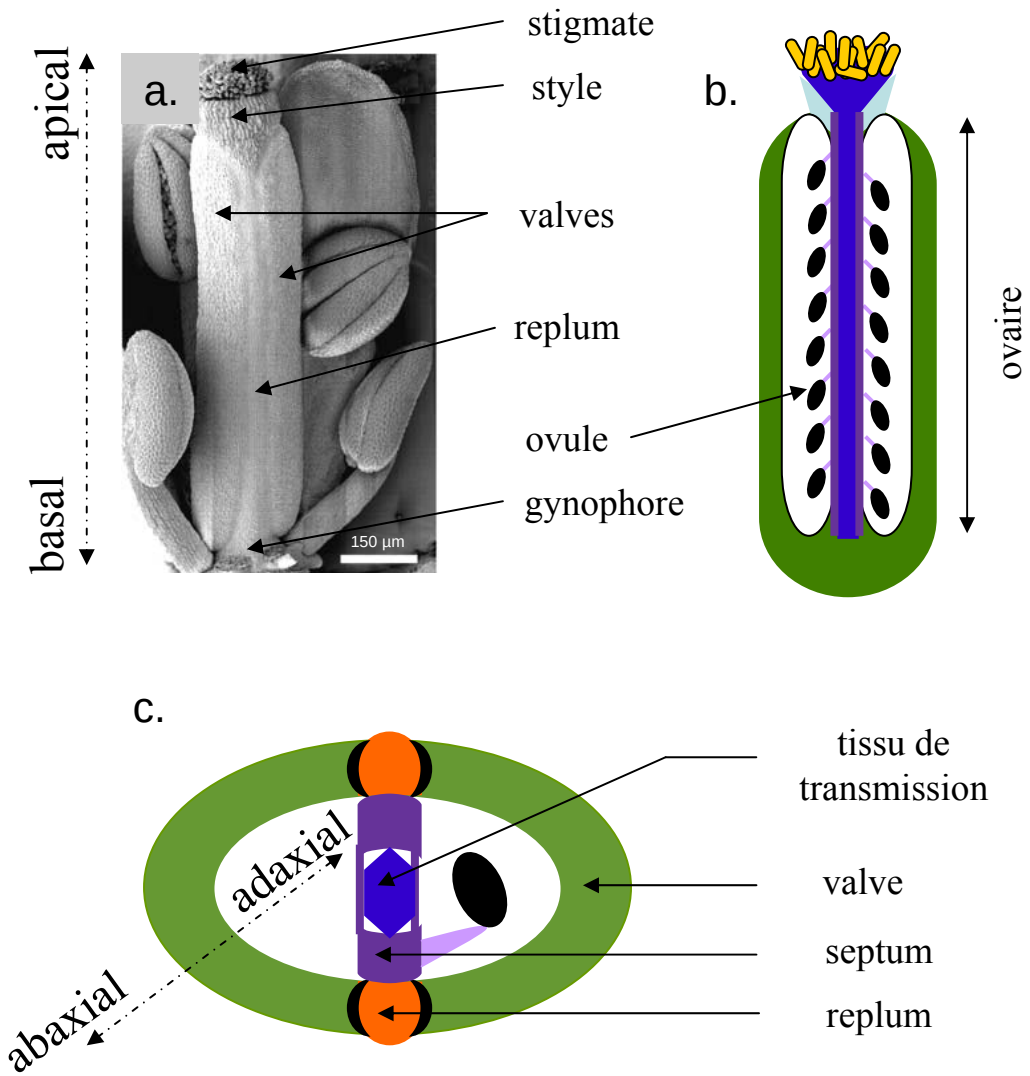


Figure 9 : Présentation du carpelle de l'espèce modèle *Arabidopsis thaliana*. (a) Cliché de microscopie électronique à balayage. (b) Coupe longitudinale schématique. (c) Coupe transversale schématique. Notons que le gynécée de cette espèce est constituée de deux carpelles fusionnés.

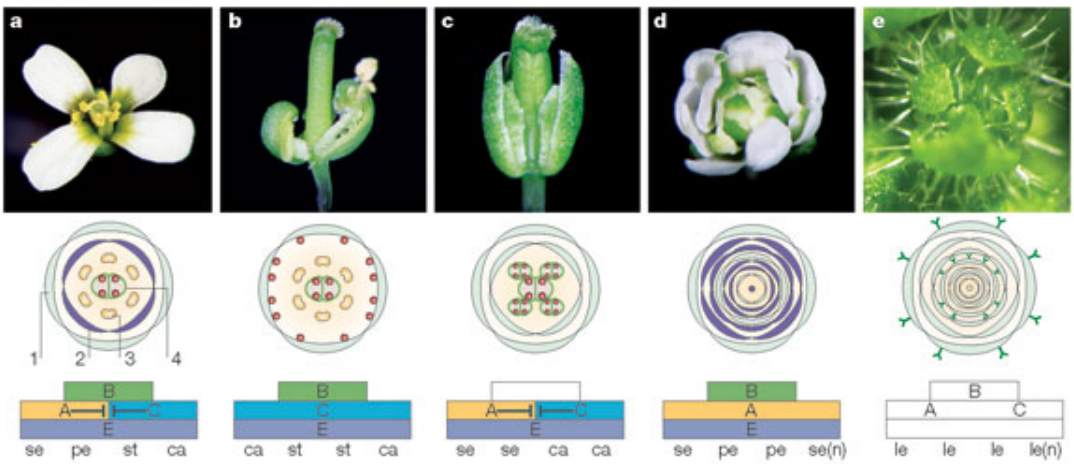


Figure 10 : Le modèle ABCE du patterning des différent organes floraux. (a) La fleur sauvage d'*Arabidopsis thaliana* consiste en quatre verticilles d'organes. L'identité des sépales (se) est conférée par l'activité des gènes de classe A dans le premier verticille, l'identité des pétales (pe) est conférée par l'activité des gènes de classe A et B dans le deuxième verticille, l'identité des étamines (st) est conférée par l'activité des gènes de classe B et C dans le troisième verticille, l'identité des carpelles (ca) est conférée par l'activité des gènes de classe C dans le quatrième verticille. L'activité de classe E est requise pour la détermination de chaque organe. (b) Phénotype du mutant *apetala2* (*ap2*) de classe A . (c) Phénotype du mutant *pistillata* (*pi*) de classe B. (d) Phénotype du mutant *agamous* (*ag*) de classe C. (e) Phénotype du quadruple mutant *sepallata1,2,3,4* (*sep1sep2sep3sep4*) de classe E. Cette figure est tirée de la revue de Krizek et Fletcher (2005) sur le développement floral. NB: on étend parfois ce modèle à ABCDE selon si l'identité des ovules est considérée indépendante ou non de la fonction C (Colombo et al., 1995).

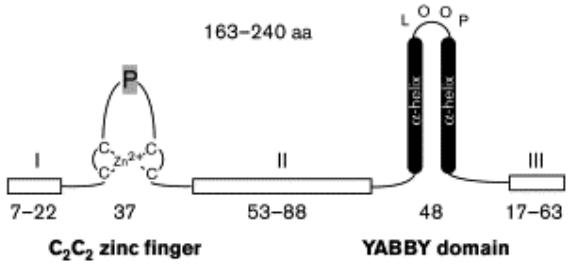
Concernant la voie indépendante d'*AG*, les gènes *SPATULA* (*SPT*) et *CRABS CLAW* (*CRC*) ont été identifiés comme des composantes de(s) voie(s) indépendante(s) d'*AG*. Pour preuve, l'inactivation des gènes *SPT* et *CRC* dans un fond génétique *ap2 ag* conduit à un phénotype similaire à celui du quadruple mutant *ap2 ag shp1 shp2* (Alvarez and Smyth, 1999).

Des homologues du gène *AG* ont été clonés chez de nombreuses espèces de spermaphytes, permettant ainsi de reconstruire l'histoire évolutive de ce gène (Kramer et al., 2004). Certains gènes *AG* hétérologues ont été exprimées de manière ectopique chez *Nicotiana tabacum* (Kang, 1995; Mandel et al., 1992) et chez *A. thaliana* (Benedito et al., 2004) afin de tester la fonction des différents gènes *AG*. Les phénotypes obtenus sont semblables à des plantes exprimant *AG* de manière constitutive chez *A. thaliana* (Mizukami and Ma, 1992). Ces résultats suggèrent que la conservation de séquence entre les homologues d'*AG* reflète une conservation de leurs interactions biochimiques (Kramer et al., 2004).

3) Polarité adaxiale/abaxiale

Chez *A. thaliana*, la famille YABBY est composée des six gènes suivants : *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER* (*FIL* ou *YAB1*), *YAB2*, *YAB3*, *INNER NO OUTER* (*INO* ou *YAB4*) et *YAB5* **(figure 11)**. Chacun de ces gènes est exprimé dans la région abaxiale d'un ou plusieurs organes latéraux. *CRC* a été le premier gène à être cloné par mutagenèse EMS lors d'un crible de mutants du carpelle. L'histoire évolutive du clade CRC commence progressivement à se dessiner au sein des angiospermes. Chez *A. thaliana*, *CRC* intervient dans la mise en place de la polarité adaxiale/abaxiale du carpelle, et se caractérise par une expression : i) restreinte au carpelle et nectaires, ii) polarisée de manière abaxiale dans la paroi carpellaire. Au contraire, *DROOPING LEAF* (*DL*), l'orthologue de *CRC* chez *Oryza sativa*, est exprimé très tôt au cours du développement du carpelle et des feuilles (Yamaguchi et al., 2004). Les mutants perte-de-fonction *dl* présentent des conversions homéotiques de carpelles en étamines, ainsi que des défauts de développement de la nervure centrale du limbe foliaire (Yamaguchi et al., 2004). Il semblerait donc que *DL* soit impliqué dans la détermination de l'identité carpellaire (et non dans la polarité abaxiale) ainsi que dans le contrôle du développement foliaire chez *O. sativa*. Récemment, la caractérisation de l'orthologue *AmbCRC* chez *A. trichopoda* a montré une forte conservation du patron d'expression et de la fonction de ce gène avec *CRC* chez *A. thaliana* (Fourquin et al., 2007 ; Fourquin et al., 2005). L'ensemble de ces données suggère que *CRC* devait probablement avoir une expression polarisée de manière abaxiale dans le carpelle de l'ancêtre commun des

(a')



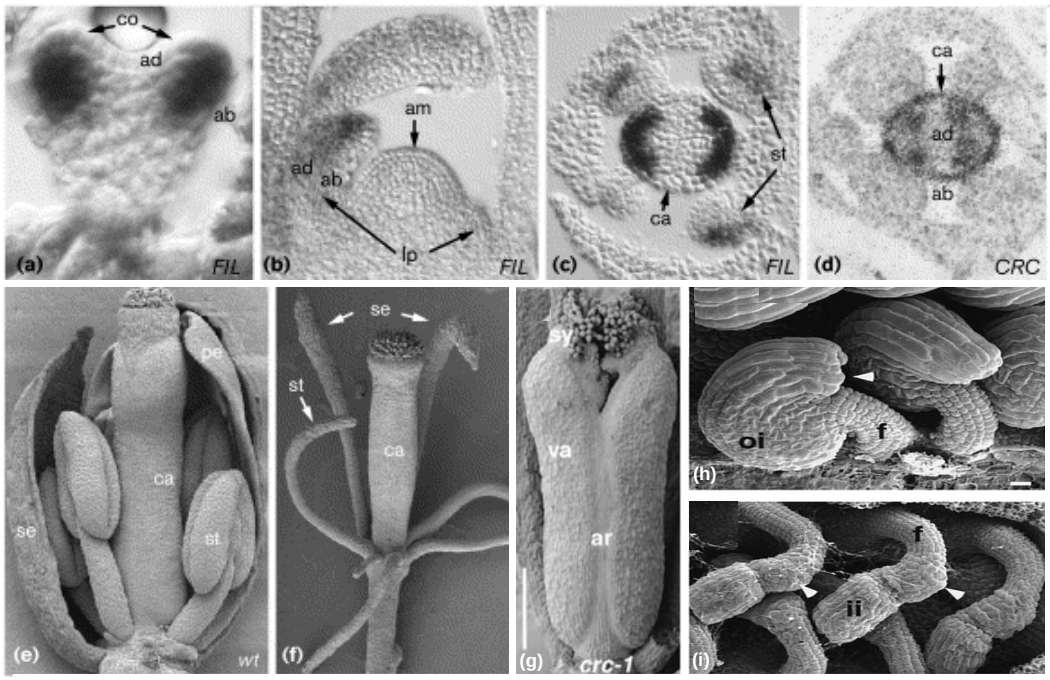


Figure 11 : Présentation générale de la famille YABBY chez *A. thaliana*. Cette famille de facteurs de transcription comprend les six membres suivants chez *A. thaliana* : *CRABS CLAW* (*CRC*), *INNER NO OUTER* (*INO*), *FILAMENTOUS FLOWER* (*FIL*), *YABBY2*, *YABBY3* et *YABBY5* (Bowman, 2000). La structure protéique consiste en un domaine zinc-finger et un domaine de type hélice-boucle-hélice appelé domaine YABBY (a'). Ces gènes ont la caractéristique commune d'avoir une expression polarisée de manière abaxiale dans les organes latéraux tels que les cotylédons (a), les primordia foliaires (b), les étamines (c), et les carpelles (c, d). Le domaine d'expression de ces gènes est indiqué par des flèches noires. Par comparaison avec une plante sauvage (e, h), le double mutant *fil yab3* présente des organes floraux radialisés (f), le mutant *crc* présente un défaut de fusion des carpelles (g), et le mutant *ino* présente des ovules dépourvus de tégument externe (i). Les défauts des phénotypes mutants sont indiqués par des flèches blanches. ab, abaxial ; ad, adaxial ; am, méristème apical ; ca, carpelle ; co, cotylédon ; f, funicule ; ii, tégument interne ; lp, primordium foliaire ; oi, tégument externe ; pe, pétale ; se, sépale ; st, étamine ; sy, style ; va, valve.

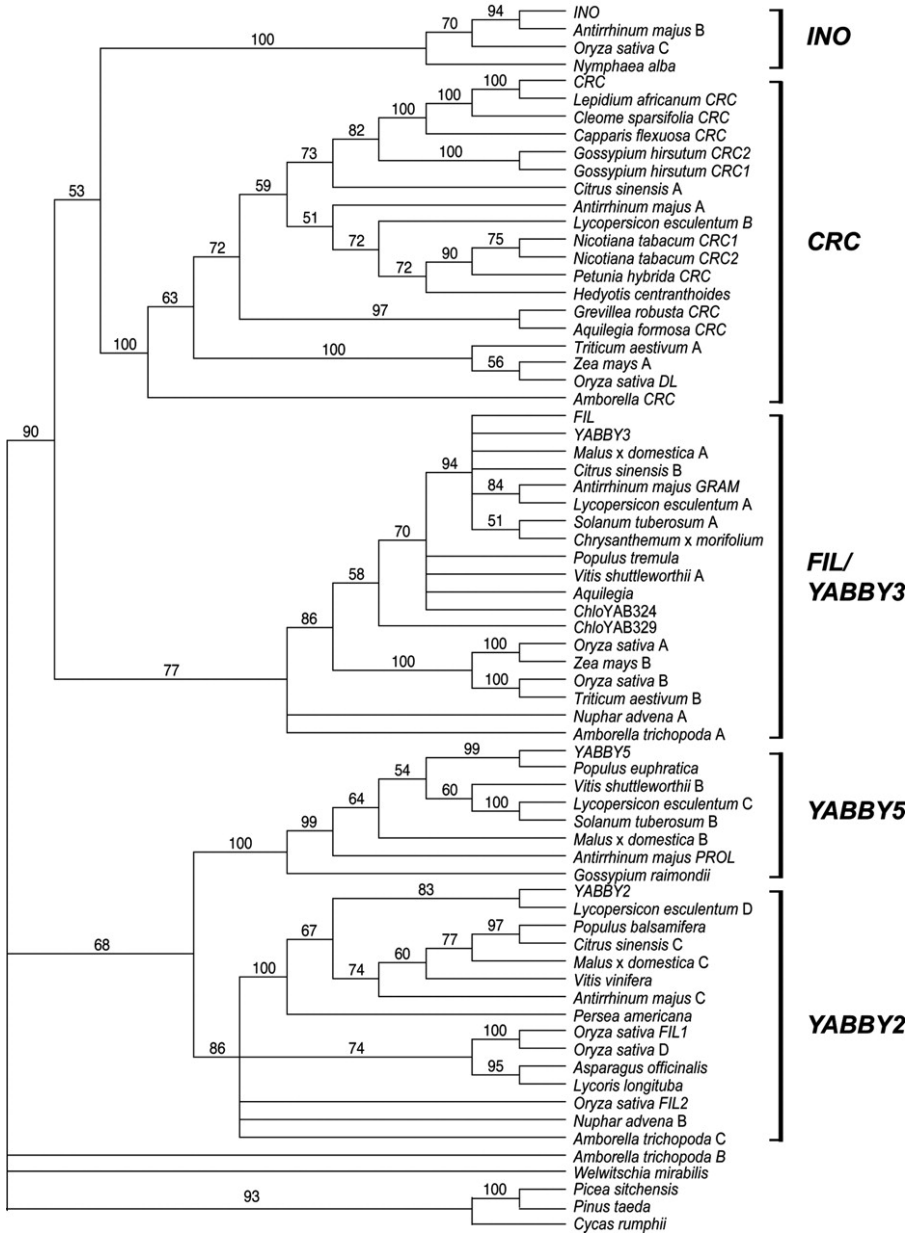


Figure 12 : Répartition phylogénétique de la polarité adaxiale-abaxiale. Le nombre croissant de séquences YABBY disponibles a permis la reconstruction phylogénétique de cette famille au sein des angiospermes (Lee et al., 2005). Les gènes YABBY se répartissant en cinq clades, il semblerait donc que l'ancêtre commun des plantes à fleurs actuelles possédait cinq gènes YABBY. Chez les eudicots, ces gènes interviennent dans la détermination du caractère abaxial de certains organes latéraux. Cette fonction ne semble pas conservée chez l'ensemble des angiospermes étant donnée l'absence de polarité chez plusieurs gènes YABBY d'*O. sativa*, ainsi que la polarité adaxiale de certains gènes chez *A. trichopoda* et *Z. mays*.

plantes à fleurs. Selon cette hypothèse, les changements d'expression et de fonction de *DL* auraient été acquis après la divergence du lignage des Poacées.

Les clades INO et YAB2 commencent également à être bien documentés. Chez *A. thaliana*, *INO* a un rôle dans le développement du tégument externe de l'ovule comme semblent l'indiquer les mutants *ino*, dépourvus de tégument externe alors que le tégument interne se développe normalement. De plus, *INO* est exprimé spécifiquement dans l'ovule chez les angiospermes avec la présence de transcrits détectés uniquement dans le tégument externe des ovules chez *A. thaliana*, ainsi que dans les deux téguments et le nucelle des ovules de *Nymphaea alba* (Yamada et al., 2003). Concernant le clade YAB2, il a été défini par rapport au gène *YAB2* d'*A. thaliana* dont la fonction n'est pas clairement établie. Ce gène est impliqué dans la détermination du caractère abaxial de certaines cellules comme l'atteste son expression polarisée dans les primordia des cotylédons et des organes floraux. Encore une fois, la mise en place du caractère abaxial de *YAB2* ne semble pas généralisable à l'ensemble des angiospermes. L'homologue *OsYAB1* chez *O. sativa* ne présente pas une expression polarisée et semble plutôt intervenir dans le maintien de l'identité "organe reproducteur" (Jang et al., 2004), alors que l'homologue *AmbF1* chez A. *trichopoda* présente une expression polarisée mais de manière adaxiale dans le carpelle et les feuilles.

Les clades FIL et YAB3 sont les moins bien documentés chez l'ensemble des angiospermes. Chez *A. thaliana*, *FIL* et *YAB3* sont deux proches paralogues qui présentent une redondance fonctionnelle. Comme la majorité des gènes YABBY chez *A. thaliana*, ces gènes interviennent dans l'établissement du caractère abaxial des organes latéraux et leur expression est donc polarisée chez les eudicots. Cependant, aucun des homologues clonés chez les monocots ne présente une telle expression polarisée, avec peut-être une exception pour un orthologue de *FIL/YAB3* exprimé de manière adaxiale chez *Zea mays* (Juarez et al., 2004).

La reconstruction phylogénétique de la famille YABBY au sein des angiospermes actuelles révèle que la plupart des lignages YABBY étaient déjà présents à la base des angiospermes **(figure 12)** (Lee et al., 2005; Toriba et al., 2007). La famille YABBY devait contenir au moins quatre membres chez l'ancêtre commun des plantes à fleurs. Cependant, les phylogénies actuelles de la famille YABBY restent incongruentes entre-elles, notamment au niveau des clades YAB2 et YAB5.

4) Polarité apico-basale

D'autres mutants sont connus pour présenter des régions apicales élargies et des ovaires réduits. Par exemple, les mutants *pinoid* (Bennett et al., 1995) et *pinformed* (Okada et al., 1991) sont affectés dans le transport de l'auxine et les mutants *ask*α et *ask*γ (Dornelas et al., 2000) possèdent des kinases incapables de réaliser une transduction parfaite du signal. L'établissement d'un gradient d'auxine constitue une hypothèse intéressante pour expliquer la polarité apico-basale du carpelle, même si cela ne reste qu'un modèle à l'heure actuelle (Balanza et al., 2006; Nemhauser et al., 2000). Cependant, des données récentes font le lien entre les gènes *STY*/*SHI* et la synthèse de l'auxine (Sohlberg et al., 2006). Les auteurs identifient *YUCCA4*, un gène codant une enzyme centrale dans la biosynthèse de l'auxine, comme une cible directe du facteur de transcription *STY1*. De plus, ils montrent que les gènes *STY1* et *STY2* sont exprimés à l'apex du carpelle en train de se développer - justement à l'endroit où la source d'auxine est supposée être – et que la concentration en auxine libre est plus faible dans les mutants *sty*/*shi*. En parallèle, des données sur la caractérisation des gènes *YUCCA* révèlent que ces gènes sont exprimés dans la région apicale du carpelle, et que les mutants présentent des phénotypes sévèrement altérés au niveau du gynécée (Cheng et al., 2006).

Pour résumer le modèle actuel, la mise en place de la polarité apico-basale repose sur la synthèse de l'auxine à l'apex du gynécée. Cette biosynthèse se fait via l'expression du gène *YUCCA* qui serait directement activé par les gènes *STY*/*SHI* lors des étapes précoces du développement carpellaire. Le gène *SPT* pourrait répondre à ces fortes concentrations d'auxine en déclenchant d'une part la différenciation du stigmate, du style et des tissus de transmission, d'autre part en participant à la formation du gradient d'auxine en régulant le transport polaire de cette hormone. Ce rôle hypothétique de *SPT* est en accord avec la complémentation partielle de mutants *spt* par traitement au NPA, un inhibiteur du transport polarisé de l'auxine (Nemhauser et al., 2000). Concernant le gène *ETTIN* (*ETT*), il pourrait répondre à des concentrations intermédiaires en auxine pour mettre en place l'ovaire. En tant que membre des facteurs de transcription ARF, ETT (ou ARF3) est capable d'interpréter les gradients d'auxine et de transmettre l'information par activation ou répression des gènes cibles **(figure 13)**.

D'une manière générale, les facteurs de transcription de la famille ARF (**A**uxin **R**esponse **F**actor) jouent un rôle central dans la réponse à l'auxine. Chez *A. thaliana*, ils constituent une famille de 23 membres numérotés de ARF1 à ARF23. La majorité des protéines ARF possèdent un domaine de liaison à l'ADN (divisé en un domaine I de type B3 que l'on

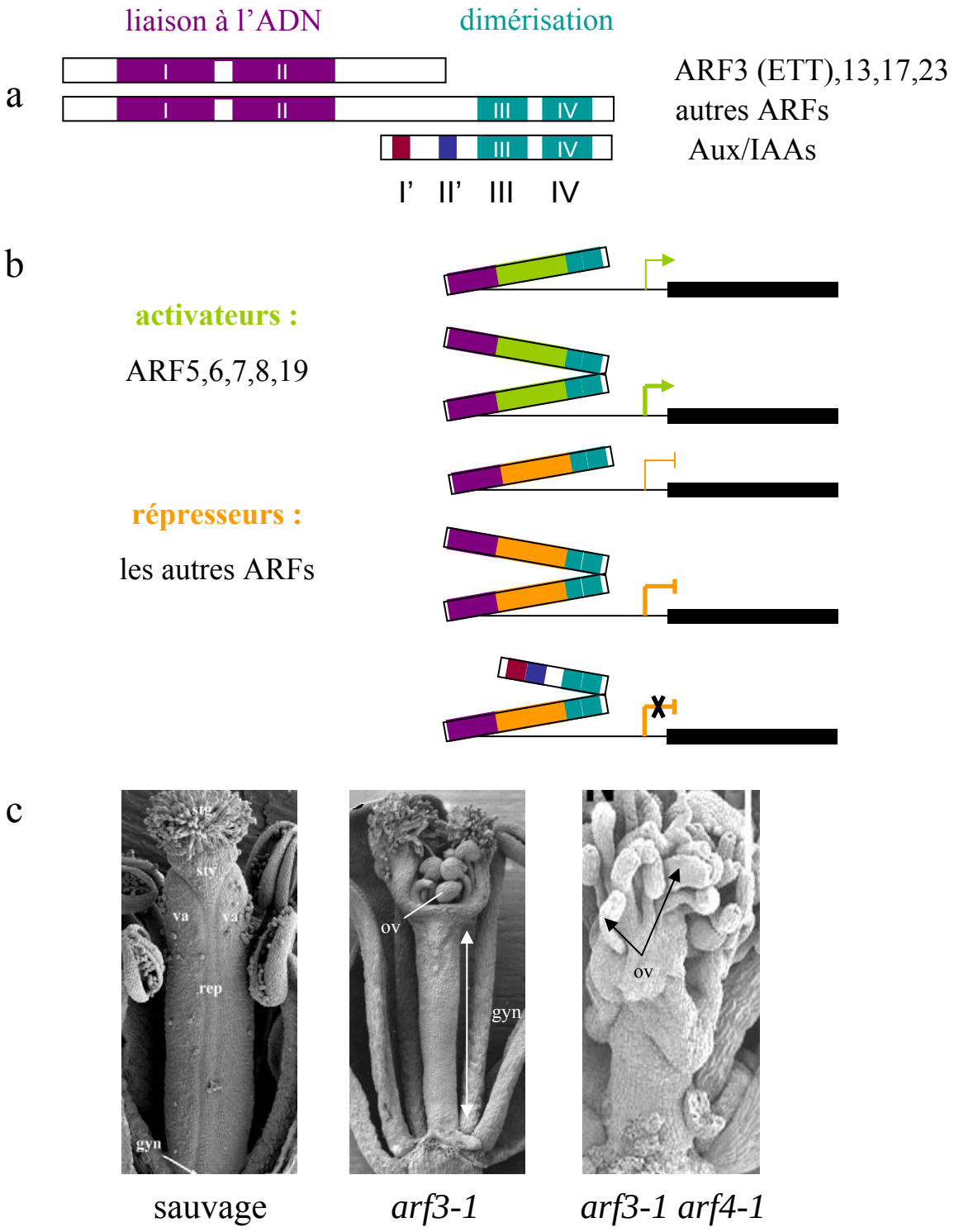


Figure 13 : Présentation générale de la famille ARF chez *A. thaliana*. A) Cette famille de facteurs de transcription comprend 23 membres chez *A. thaliana*, caractérisés par un domaine de liaison à l'ADN et un domaine d'interaction protéine-protéine (sauf ARF3,13,17,23). **B)** Selon la nature de la zone centrale, les ARFs sont des activateurs ou des répresseurs transcriptionnels. **C)** Phénotype des mutants *arf3* et *arf3 arf4* présentant des défauts au niveau du carpelle. gyn, gynophore ; ov, ovule ; rep, replum ; stg, stigmate ; sty, style ; va, valve.

retrouve dans d'autres classes de facteurs de transcription + un domaine II ou domaine ARF spécifique de la famille) ainsi que des domaines III et IV impliqués dans les interactions avec d'autres protéines. Ces interactions peuvent être de natures différentes : (i) homodimérisation entre ARFs, (ii) hétérodimérisation entre ARFs, (iii) interaction avec des régulateurs négatifs de type Aux/IAA. Chez *A. thaliana*, la famille Aux/IAA comprend 29 membres caractérisés par la présence de domaines III et IV qui interviennent dans la liaison avec les domaines III et IV des ARFs. Notons ici que les domaines III et IV présents chez les deux familles protéiques ont toujours été considérés comme homologues sans la moindre preuve.

De manière intéressante, les protéines ARF3, ARF13, ARF17 et ARF23 diffèrent en structure par une absence des domaines III et IV **(figure 13a)**. Or, nous savons que le gène *ARF3* est central dans le développement du carpelle et qu'il interagit génétiquement avec *ARF4*, son paralogue le plus proche chez *A. thaliana* **(figure 13c)**.

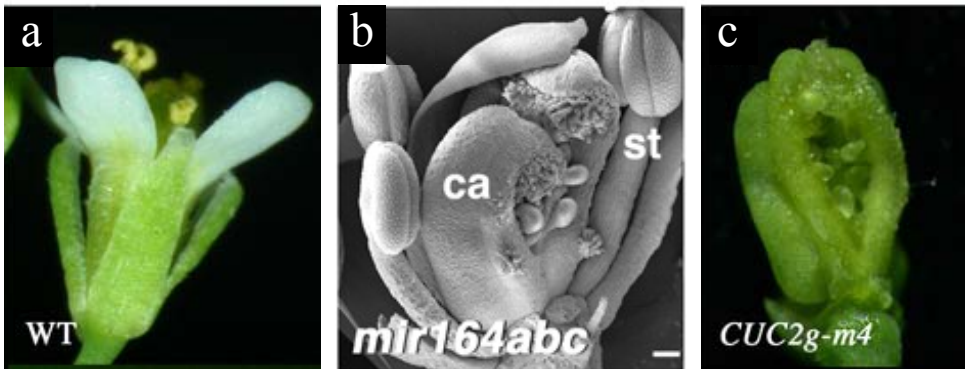
5) Fusion des carpelles

Chez *A. thaliana*, la famille génique *MIR164* se compose des gènes *MIR164a*, *MIR164b* et *MIR164c* qui ciblent plusieurs membres de la famille NAC dont les gènes *CUP-SHAPED COTYLEDON1* (*CUC1*) et 2 (*CUC2*). Le mutant *MIR164c* présente des défauts de fusion au niveau du gynécée (Baker et al., 2005), alors que le triple mutant *MIR164abc* se caractérise par une absence totale de fusion des carpelles **(figure 14b)** (Sieber et al., 2007). De la même manière, la fusion des carpelles est altérée chez des plantes exprimant une version de *CUC2* muté dans le domaine cible de *miR164* (Peaucelle et al., 2007).

IV) Présentation de la problématique

1) Introduction

L'apparition soudaine et massive d'une grande diversité de plantes à fleurs (ou angiospermes) dans les gisements fossiles du Crétacé inférieur est considérée, depuis plus d'un siècle, comme un "abominable mystère". Cette expression, utilisée par Darwin en 1879 dans une lettre au botaniste Sir Joseph Hooker, est encore d'actualité malgré les efforts déployés pour tenter d'élucider cette énigme. En effet, les angiospermes auraient évolué à partir d'un ancêtre apparenté aux gymnospermes il y a 150 Ma, et se seraient rapidement diversifiées pour donner les 300 000 espèces actuelles. La reconstruction de la phylogénie moléculaire des plantes à graines a abouti à plusieurs résultats déterminants. Tout d'abord, les angiospermes et les gymnospermes regroupent l'ensemble des plantes à graines et représentent deux groupes monophylétiques qui auraient divergé il y a 300 Ma. De plus, il est



balance *CUC* / *miR164*

précurseur *miR164a*

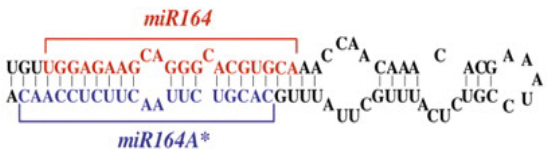


Figure 14 : Présentation générale des gènes *MIR164* et *CUC* chez *A. thaliana*. Les transcrits des gènes *CUC1* et *CUC2* sont négativement régulés par le microARN miR164 qui entraîne indirectement leur dégradation. La fonction développementale de *miR164* a pu être mise en évidence soit par mutation des gènes *MIR164a,b,c* codant pour *miR164* (b), soit par génération de formes de *CUC1,2* insensibles à *miR164* (c). Pour comparaison, le phénotype sauvage est indiqué en (a).

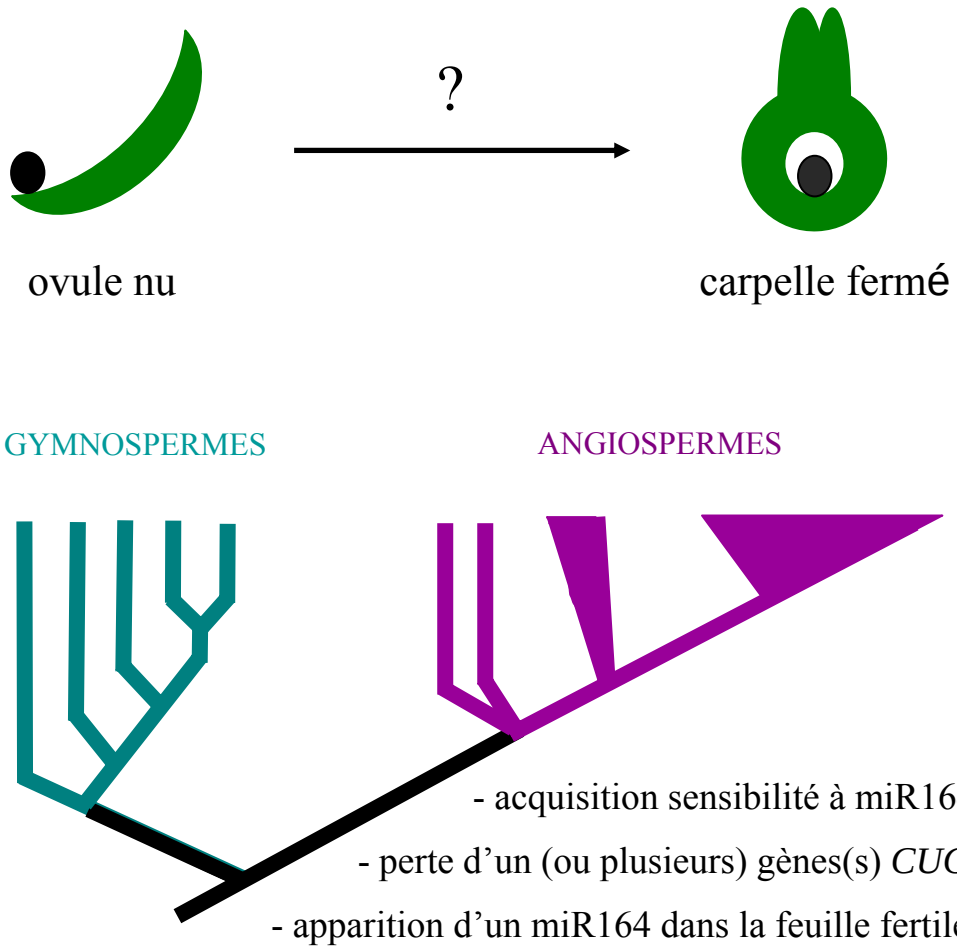


Figure 15 : Hypothèses évolutives reliant fermeture du carpelle et baisse du niveau d'expression des gènes *CUC*. Les études fonctionnelles chez *A. thaliana* montrent qu'une augmentation du niveau d'expression du gène *CUC2* conduit à l'ouverture des carpelles. Ainsi, il est logique de faire l'hypothèse qu'une diminution du niveau d'expression des homologues de *CUC2* dans le lignage pré-angiosperme aurait pu conduire à la fermeture de la feuille fertile. On peut envisager plusieurs mécanismes non exclusifs comme : (i) l'acquisition de la sensibilité à *miR164* sachant qu'un gène *MIR164* existe chez *Pinus sp*. Sans aucune validation fonctionnelle, (ii) perte d'un (ou plusieurs) gène(s) *CUC*, (iii) apparition d'un *miR164* dans la feuille fertile.

possible de distinguer des lignages ayant divergé très tôt au cours de l'évolution des angiospermes et qui sont regroupés dans le grade ANA.

L'apparition de la fleur a nécessité trois événements majeurs, dont la chronologie reste indéterminée : (i) le regroupement des organes reproducteurs mâles et femelles sur un même axe, (ii) l'internalisation des ovules au sein d'un tissu : le carpelle, (iii) la mise en place de pièces stériles en périphérie des organes reproducteurs.

Durant ma thèse, j'ai identifié certains des événements moléculaires à l'origine du carpelle des plantes à fleurs. Dans un premier temps, j'ai comparé les mécanismes moléculaires du développement carpellaire entre la plante modèle *Arabidopsis thaliana* et des angiospermes basales du grade ANA. Cette comparaison m'a permis d'inférer des données quant au développement de la fleur ancestrale. Dans un second temps, j'ai étudié le développement reproducteur chez les gymnospermes actuelles que j'ai par la suite comparé à celui de notre modèle chez la fleur ancestrale. Ainsi ont été mis en évidence des changements moléculaires responsables de l'apparition de la fleur.

2) Vers un "portrait moléculaire" de la fleur ancestrale

Trois angiospermes basales du grade ANA, *Amborella trichopoda* (Amborellaceae), *Cabomba aquatica* (Nympheaceae) et *Illicium parviflorum* (Illiciaceae) ont été choisies afin de construire un modèle de la morphogenèse florale chez l'ancêtre commun des plantes à fleurs. En parallèle a été dressée une liste de gènes connus pour intervenir dans le développement floral chez *A. thaliana*, et dont il me paraissait intéressant de retracer l'histoire évolutive. En particulier, j'ai opté pour l'étude des gènes des familles YABBY et ARF.

Ainsi, nous avons incorporé ces gènes dans un modèle moléculaire de la morphogenèse florale chez la fleur ancestrale, selon l'approche suivante. Tout d'abord une étape qui consistait en le clonage des gènes d'intérêt (ou plutôt les ADNc correspondants) chez les espèces du grade ANA. Ensuite, la reconstruction phylogénétique retraçant l'histoire évolutive de ces gènes permettait de : (i) s'assurer de la relation d'orthologie et non d'une simple relation d'homologie avec les séquences d'*A. thaliana* ; (ii) identifier certains des événements moléculaires qui ont mené à la phylogénie actuelle de nos séquences. Concernant les vrais orthologues, leurs patrons d'expression étaient étudiés par hybridation *in situ* et comparés à ceux qui sont établis chez *A. thaliana*. En effet, lorsque le patron d'expression est identique entre une espèce du grade ANA et *A. thaliana*, il est possible de conclure à un patron d'expression similaire chez la fleur ancestrale. Une telle conservation de l'expression spatiale d'un gène indique probablement la conservation de sa fonction depuis l'ancêtre

commun des plantes à fleurs. Suivant cette démarche, il a été possible d'apporter de nouveaux éléments quant aux mécanismes de la morphogenèse florale chez la première plante à fleur. Cela a notamment été le cas lors d'une étude au sein de notre groupe pour les gènes *CRABS CLAW* (*CRC*) et *TOUSLED* (*TSL*) qui interviennent dans le développement des organes reproducteurs femelles de la fleur (Fourquin et al., 2005).

En résumé, les résultats obtenus ont été ajoutés à ceux déjà publiés dans l'idée de dresser un "portrait moléculaire" de la fleur ancestrale. Ce travail, novateur en biologie, a représenté le premier objectif de ma thèse.

3) Résoudre l'"abominable mystère"

Deux principales théories sont actuellement avancées pour rendre compte de l'apparition d'une fleur bisexuée à partir d'une structure unisexuée de type gymnosperme. La théorie "Mostly Male" propose que la fleur bisexuée dérive d'une structure ancestrale mâle sur laquelle seraient apparus des ovules de manière ectopique. Alternativement, la théorie "Out of Male/Out of Female" suggère que la bisexualité résulterait du changement de sexe de quelques organes le long de l'axe reproducteur ancestral unisexué, ceci par modification des frontières de l'expression de certains gènes. De telles théories conduisent à des hypothèses totalement différentes quant à la conservation de l'expression des gènes spécifiques du sexe entre les angiospermes et les gymnospermes. En particulier, la théorie "Mostly Male" prédit que le carpelle des plantes à fleurs est homologue aux structures reproductrices mâles des gymnospermes. La validité de ces hypothèses n'a pas encore été réellement testée. Afin de déterminer les changements moléculaires à l'origine de la fleur, il est nécessaire de retracer l'histoire de la morphogenèse florale à partir de la période précédant l'ancêtre commun des angiospermes. Cette approche a constitué le deuxième objectif de ma thèse. Pour cela, j'ai principalement focalisé mon étude sur trois espèces de gymnospermes dont l'ancêtre commun avec le lignage des plantes à fleurs précède l'apparition de la fleur : la première espèce utilisée est *Ephedra distachya* (Ephedraceae), dont les structures reproductrices présentent des similarités avec les fleurs ; la seconde espèce est *Ginkgo biloba* (Ginkgoaceae), seul représentant d'un lignage apparu il y a plus de 200 Ma ; la troisième espèce est *Cycas thouarsii* (Cycadaceae) dont le lignage a divergé très tôt au cours de l'évolution des gymnospermes. Les individus mâles et femelles de ces gymnospermes modèles sont disponibles au Jardin botanique de Lyon, avec lequel notre groupe de recherche collabore de manière étroite.

Concernant les outils disponibles chez ces gymnospermes, des banques d'ADNc ont été récemment préparées au sein de notre groupe. Elles ont été réalisées à partir de tissus reproducteurs mâles et femelles. Ces banques ont été utiles pour l'identification de gènes orthologues à ceux dont le rôle chez la fleur ancestrale a été prouvé. Le criblage des banques d'ADNc des gymnospermes a été de nouveau suivi par des analyses phylogénétiques. Lorsque la relation d'orthologie est simple (un gène d'angiosperme regroupant avec un unique gène de gymnosperme), il a été possible de conclure que l'origine de tels gènes a précédé l'ancêtre commun des plantes à graines et que, par conséquent, elle n'a pas été directement responsable de l'évolution de la fleur. Au contraire, la phylogénie groupe parfois un gène de gymnospermes avec plusieurs gènes d'angiospermes. Dans ce cas, des événements de duplication auraient pu contribuer à l'origine de la fleur notamment par acquisition de nouvelles fonctions.

RÉSULTATS

Chapitre 2 : GÈNES *ARF3-ARF4* ET ÉVOLUTION DU CARPELLE

I) Introduction et résumé de l'article

Le but de cet article est de retracer l'histoire évolutive des gènes *ARF3* et *ARF4* impliqués dans le développement du carpelle chez *A. thaliana*. Le facteur de transcription ARF3 (ou ETTIN) diffère en structure des autres membres de la famille ARF par le fait qu'il ne possède pas les domaines régulateurs III et IV intervenant dans l'interaction avec divers régulateurs transcriptionnels tels que les AUX/IAAs. De manière surprenante, son plus proche paralogue ARF4, ne présente pas cette troncation au niveau protéique.

Grâce à cette analyse, nous avons montré que chaque espèce de plante à fleur étudiée possède un transcrit tronqué (sans domaines III et IV) soit du gène *ETT*, soit du gène *ARF4*. La troncation affecte le lignage génique ARF4 chez les espèces du grade ANA alors qu'elle est présente dans le lignage ETT chez les Poales et les eudicots. La même approche réalisée chez les gymnospermes conduit à la conclusion qu'un seul homologue *ARF3/4* existe chez le groupe frère des plantes à fleurs. Ces données indiquent qu'une duplication génique a probablement eu lieu dans le lignage menant aux angiospermes actuelles. De manière encore plus intéressante, les troncations présentes chez les angiospermes ne résultent pas des mêmes mécanismes moléculaires. Par exemple, les formes tronquées du gène *ARF4* chez *C. aquatica* et du gène *ETT* chez *A. thaliana* résultent d'une troncation au niveau génique, alors que celles du gène *ARF4* chez *A. trichopoda* sont le résultat d'un épissage alternatif. Suite à ces travaux, nous avons montré que la présence d'une forme tronquée du gène *ETT* est indispensable au développement correct du carpelle chez *A. thaliana*, suggérant que les troncations auraient pu jouer un rôle important dans la mise en place d'un carpelle au cours de l'évolution.

J'ai contribué à ce travail en réalisant la majorité du travail expérimental. J'ai notamment cloné les orthologues des gènes *ETT* et *ARF4* chez *Amborella trichopoda* et *Ephedra distachya* par criblage de banques phagiques d'ADNc. Dans un deuxième temps, j'ai cloné les séquences géniques correspondant aux différents transcrits afin d'en déduire leur structure intron/exon, me permettant ainsi de mettre en évidence le mécanisme d'épissage alternatif du gène *ARF4* chez *A. trichopoda*. En parallèle, j'ai réalisé toute les reconstructions phylogénétiques présentes dans le manuscrit. J'ai réalisé la totalité des expériences d'hybridation *in situ* chez *A. trichopoda*, *C. aquatica* et *E. distachya*. Enfin, j'ai effectué environ la moitié des expériences de complémentation du mutant *ett-1* chez *A. thaliana*.

II) Article

Classification: BIOLOGICAL SCIENCES, Evolution.

How ETTIN lost its tail: a role for protein truncation in the origin of the flower.

Authors

Cédric Finet, Chloé Fourquin, Marion Vinauger-Douard, Annick Berne-Dedieu, Pierre Chambrier and Charles P. Scutt*

Address

Laboratoire de Reproduction et Développement des Plantes,
Université de Lyon,
(CNRS UMR 5667, INRA, ENS-Lyon, UCB-Lyon),
46 allée d'Italie,
69364 Lyon Cedex 07, France.

***Author for correspondence**

Tel: +33 4 72 72 86 03
Fax: +33 4 72 72 86 00
E-mail: Charlie.Scutt@ens-lyon.fr

Abstract

The genes *ETTIN (ETT)* and *ARF4* encode partially redundant, paralogous transcription factors of the Auxin Response Factor (ARF) family in *Arabidopsis thaliana*. ETT plays an essential role in carpel development, and presents an atypical structure as it lacks two regulatory domains at its C-terminus, termed domains III and IV, which are present in most other ARF proteins. We have investigated a possible role for the origin of the *ETT* and *ARF4* genes, and for genetic truncations that removed domains III and IV from the proteins they encode, in the origin of the carpel, the female reproductive organ of the angiosperms. We show that *ETT* and *ARF4* arose through a gene duplication event that occurred prior to the last common ancestor of the extant angiosperms, and that truncated *ETT* or *ARF4* transcripts are expressed in the carpels of three distinct plant lineages that diverged in the early stages of angiosperm evolution. We explain how the production of truncated ARF proteins could be a pleisiomorphic feature of the angiosperms, even though neither *ETT* nor *ARF4* can have been permanently truncated in the last common ancestor of this group. We postulate that the production of ARF4 protein molecules which lacked domains III and IV arose in an ancestor of the angiosperms through an alternative splicing mechanism that is still present in Amborellales, the most basal extant lineage of this group. This molecular feature subsequently became fixed by parallel genetic truncations in two other early-diverging angiosperm lineages. The alternative splicing mechanism that we have discovered may constitute one of the molecular events that led to the origin the closed carpel, and hence to the evolution of the flowering plants.

Introduction

The gynoecium in the *ettin-1* (*ett-1*) mutant of *Arabidopsis thaliana* consists of a radically shortened ovary and style, supported by an elongated stalk, and topped by an enlarged stigmatic tissue (1). A defect in carpel fusion is also present in the *ett-1* gynoecium, which is particularly apparent at the apex of this structure, in which the stigmatic tissue is typically divided in two. Furthermore, the gynoecium of the *ett-1* mutant frequently shows defects in abaxial-adaxial polarity- a form of polarity which defines the orientation of plant lateral organs relative to the growth axis (adaxial = towards the axis, abaxial = away from the axis). In the case of the *ett-1* gynoecium, such defects lead to the ectopic production of transmitting tissues in the abaxial (external) domain of the gynoecium (1).

The *ETT* locus (2) encodes a transcription factor of the Auxin Response Factor (ARF) family, which includes 23 members in *A. thaliana* (3). Hence, *ETT* is also known as *ARF3*. ARF proteins are known to mediate the expression of genes containing Auxin Response Elements (AREs) in their promoters, many of which are induced in response to auxin (4). ARFs typically contain four conserved protein domains: a B3-like domain and an ARF domain at the N-terminal end, which together function in DNA binding, and domains III and IV near the C-terminus, which function in protein-protein interactions (3). The domains III and IV of ARF proteins are named for their homology to similar domains in Aux/IAA proteins (5). These latter proteins are able to negatively regulate of ARFs by the formation of heterodimers through the domains III and IV of both classes of molecule (6, 7). Aux/IAA proteins are themselves negatively regulated by a small family of F-box proteins, including TIR1, which can target Aux/IAA proteins for destruction in proteasomes in response to a direct interaction with auxin (8, 9). Thus, ARFs appear to be activated by auxin, via the auxin-dependent destruction of Aux/IAA proteins. The domains III and IV of ARF proteins are also known to facilitate homodimerization, or heterodimerization between different ARFs, which is required in some ARF proteins for *in vitro* binding to AREs (10).

ETT is an atypical member of the ARF protein family as it truncated at its C-terminus and consequently does not contain domains III and IV (2). Hence, ETT can be assumed to be insensitive to direct negative regulation by Aux/IAA proteins, and also to be capable of binding to cis-acting elements in its target gene promoters without interactions with ARF proteins via domains III and IV. By contrast, *ARF4*, the closest paralog to *ETT* in the *A. thaliana* genome, encodes a protein of typical ARF structure, including domains III and IV

(11). Plants in which *ARF4* is inactivated show no apparent mutant phenotype. However, *ett/arf4* double mutants show a general breakdown in the adaxial-abaxial polarity of lateral organs and an increased severity of gynoecium phenotype compared to *ett* single mutants (12).

The redundant control of adaxial-abaxial polarity by *ETT* and *ARF4* may occur partially as a result of the adaxial expression of two *TAS3*-derived transactivating, small interfering RNAs (ta-siRNAs) that target *ETT* and *ARF4* (13). In addition to post-transcriptional regulation by ta-siRNAs, *ETT* is regulated at the translational level through an upstream open reading frame (uORF) in its 5'-leader sequence (14). Accordingly, the correct expression of *ETT* depends on the activity of a ribosomal protein which is involved in the translational reinitiation of polycistronic genes. Though this mechanism has only been studied in detail in *ETT* and *MONOPTEROS (MP, ARF5)* of the ARF family, *ARF4* also contains uORFs in its 5'-leader sequence, as do approximately half of the members of *A. thaliana* ARF family (15), suggesting this form of regulation may be widespread among ARF proteins.

Molecular phylogenetic investigations over the last decade have largely resolved the relationships of the major seed plant groups. Accordingly, the extant seed plants appear to be made up of two clades corresponding to the angiosperms (the flowering plants) and the gymnosperms (non-flowering seed plants), respectively. The last common ancestor shared by these two groups is estimated to have lived some 300 MYA (million years ago) (16, 17). By contrast, molecular clock estimates have dated the last common ancestor of the living angiosperms to only around 160 MYA (18), a date which is reasonably consistent with the earliest known angiosperm fossils from the Lower Cretaceous, some 130 MYA. The flowering plant lineage, therefore, appears to have separated from that of the living gymnosperms some 140 MYA before the origin of the flower.

Within the angiosperm clade, three lineages with living representatives, Amborellales, Nymphaeales and Austrobaileyales, collectively known as the ANA grade, appear to have diverged at an early stage from a remaining common lineage that is shared by all other living flowering plants (19). Evidence from INDEL changes (20, 21) confirms the finding of earlier phylogenetic analyses that Amborellales and Nymphaeales diverged from the remaining common lineage before Austrobaileyales. The relative order of divergence of Amborellales and Nymphaeales remains to be fully resolved (22, 23), though most recent analyses place

Amborellales alone as sister to all other angiosperms (18, 24). Amborellales is represented by the single living species of *Amborella trichopoda*, a shrub that is endemic to the tropical rainforests of New Caledonia. Nymphaeales, by contrast, is a widely distributed order of aquatic plants, comprising the three families of Nymphaeceae, Cabombaceae and Hydatellaceae (25, 26), which contain a total of nine genera. The comparison of ANA grade angiosperms can be used to draw conclusions on the last common ancestor of the extant angiosperms, while comparisons with their more distant relatives, the living gymnosperms, may be used to provide insight into changes that led to the evolution of the flower, and other unique features of the flowering plants.

One of the major novel features of the flowering plants is the carpel. As *ETT* is crucial to carpel development in *A. thaliana*, and also presents an atypical, truncated structure within the ARF family, we undertook to discover the stage of evolution at which the *ETT* gene arose as a distinct ARF lineage, and that at which it became truncated. Here, we present the results of this investigation, which have permitted us to construct a hypothesis for the role of truncated ARF transcripts in the origin of the carpel- the defining morphological feature of the angiosperms.

Results

At least three distinct mechanisms for the production of ETT and ARF4 proteins that lack regulatory domains III and IV have arisen during angiosperm evolution.

To investigate the potential role of the *ETT* gene and its molecular truncation in the origin of the carpel, we decided to determine the stage of evolution at which the *ETT* lineage diverged from that of its paralog *ARF4*, and also the stage at which *ETT* became truncated. Accordingly, we identified cDNAs that were homologous to *ETT* and/or *ARF4* in two ANA grade angiosperms, *Amborella trichopoda* (Amborellales, Amborellaceae) and *Cabomba aquatica* (Nymphaeales, Cabombaceae), whose lineages diverged from a remaining common lineage at the base of the tree of the extant flowering plants, and in the gymnosperm *Ephedra distachya* (Gnetales, Ephedraceae), whose lineage separated from that of the angiosperms before the origin of the flower. We analyzed the phylogenetic relationships of the novel sequences identified, together with those of putative *ETT* and *ARF4* orthologs from the angiosperms *A. thaliana, Oryza sativa, Populus trichocarpa*, *Cucumis sativa* and *Vitis vinifera*, for which full genome sequence data were available (Fig 1A). Sequences from all of the angiosperm species sampled were resolved in our analysis into two well-supported clades,

respectively containing the *ETT* and *ARF4* genes of *A. thaliana* (Fig 1A). Both of these clades included, in basal positions, genes from the ANA grade angiosperms *A. trichopoda* and *C. aquatica*. By contrast, *Edi-ARF3/4*, from the gymnosperm *E. distachya*, occupied a sister position to the combined angiosperm *ETT* and *ARF4* clades (Fig 1A). These data strongly suggest the duplication that generated the *ETT* and *ARF4* lineages to have occurred after the divergence of the extant angiosperms and gymnosperms, but before the last common ancestor of the extant angiosperms. *ARF4* orthologs are absent from *O. sativa*, and have not yet been found in any monocot species, suggesting that *ARF4* was lost from the monocot lineage, after its separation from that of the eudicots.

The putative *ETT* orthologs identified from *A. trichopoda* and *C. aquatica* were found to encode proteins that possessed domains III and IV, whereas the putative *ARF4* orthologs from these species were found to encode proteins that lacked these domains (Fig 1B). Thus, the structure of putative *ETT* and *ARF4* orthologs in two ANA grade angiosperms appears to be the converse of the situation in *A. thaliana*, other eudicots and monocots. These surprising results prompted us to examine the genomic loci corresponding to the novel ARF cDNAs identified in ANA grade angiosperms. Typical ARF genes contain 11 introns in well conserved positions, the last two of which border sequences encoding protein domains III and IV. When aligned with their respective mRNAs, the loci encoding putative *ETT* orthologs from both *A. trichopoda* and *C. aquatica* identified the 11 introns of typical ARF genes (Fig 1B). However, the putative ARF4 locus from *C. aquatica*, *Caq-ARF4*, lacked the eleventh of these conserved introns (Fig 1B). This observation clearly indicated the truncation of *Caq-ARF4* to be distinct from that of *ETT* orthologs in *A. thaliana*, other eudicots and monocots, which lack both the tenth and eleventh introns of typical ARF genes (Fig 1B). Initial alignment of the genomic sequence of *Atr-ARF4* from *A. trichopoda* with that of its mRNA identified ten introns. However, closer inspection of the tenth of these introns revealed the presence of sequences encoding domains III and IV, which were spliced out of the mature transcript. This observation prompted us to perform Reverse Transcriptase PCR (RT-PCR) analyses on *A. trichopoda* female flower RNA using primers designed from the *Atr-ARF4* coding sequence. This procedure revealed two transcripts of different lengths (data not presented), representing the alternative splicing event shown in Fig 1B, and indicating that eleven introns are present in *Atr-ARF4* with respect to its longer transcript. The predominant RT-PCR product amplified in this procedure corresponded to the shorter transcript shown in Fig 1B, which encoded a protein lacking domains III and IV. This transcript was also much

more highly represented in an *A. trichopoda* female flower cDNA library than was the longer transcript (data not presented), again suggesting the shorter form to represent the predominant mature *Atr-ARF4* transcript *in vivo.* We checked for similar alternative splicing events in *ARF4* from *A. thaliana* and in *Caq-ETT* from *C. aquatica*, both of which encode proteins containing domains III and IV. However, no such alternative transcripts could be found for these genes. Hence, alternative splicing to generate an ARF transcript that lacked domains III and IV seemed to be restricted, among the species analyzed, to *A. trichopoda*.

Gene expression patterns in early-diverging lineages are consistent with a possible role for ARF protein truncations in carpel evolution.

In *A. thaliana*, both *ETT* and *ARF4* are expressed in all above-ground lateral organs, including carpels and other floral organs. To assess the conservation of expression of these genes in widely diverged seed plant lineages, we analyzed the expression of their orthologs in reproductive and vegetative tissues of ANA grade angiosperm and gymnosperm species. Northern blot hybridizations revealed comparable expression levels of *ETT* and *ARF4* orthologs in flower and leaf tissues of the ANA grade angiosperms *A. trichopoda* and *C. aquatica* (Fig 2A), which was consistent with the expression patterns of *ETT* and *ARF4* in *A. thaliana*. Similarly, *Edi-ARF3/4* transcripts accumulated to approximately equivalent levels in male and female cones, and in vegetative stems, of the gymnosperm *E. distachya* (Fig 2A).

To more precisely determine the expression patterns of *ETT* and *ARF4* orthologs in female reproductive tissues, we performed *in situ* hybridization analyses. These analyses revealed high levels of expression of *Atr-ETT* (Fig 2B), *Atr-ARF4* (Fig 2C), *Caq-ARF4* (Fig 2E) and *Edi-ARF3/4* (Fig 2F) in ovule tissues of their native species, consistent with the expression of *ETT* (Fig 2G) and *ARF4* in the ovules of *A. thaliana*. By contrast, *Caq-ETT* did not appear to be highly expressed in the ovules of *Cabomba aquatica* (Fig 2D). In addition to strong expression in ovules, *Atr-ETT, Atr-ARF4, Caq-ARF4* and *Edi-ARF3/4* were all measurably expressed in the surrounding reproductive tissues, corresponding to the ovary wall in ANA grade angiosperms (Fig 2B, C and E), and to the inner female bract, or chlamys, in *E. distachya* (Fig 2F). Expression of *ETT* in the ovary wall of *A. thaliana* shows a clear polarity, being stronger in the abaxial domain (2) (Fig 2G). The expression of ARF genes in the ovary wall of ANA grade angiosperms (Fig 2B-E), and in the chlamys of *E. distachya* (Fig 2F), appeared, however, to be less clearly polarized.

We also examined the conservation of sites associated with the post-transciptional and translational regulation of *ETT* and *ARF4* orthologs from ANA grade angiosperm and gymnosperm species. Putative target sites of the *TAS3*-derived ta-siRNAs *ta-siR2141* and *ta-siR2142* were very highly conserved in all five ARF genes analyzed from ANA grade angiosperm and gymnosperm species (Fig 2H). Similarly, uORFs were found in the 5'-leader sequences of *Caq-ETT*, *Atr-ARF4* and *Edi-ATF3/4* (Fig 2I). These results strongly suggest mechanisms for both the post-transcriptional and translation regulation of genes of the *ETT/ARF4* clade to have been conserved in the angiosperm and gymnosperm lineages since the last common ancestor of the seed plants.

Taken together, the above data suggest the expression patterns of *ETT* and *ARF4* orthologs in female reproductive tissues, and certain mechanisms regulating their expression, to be relatively well-conserved in the angiosperm lineage since considerably before the origin of the flower. In particular, at least one ARF transcript encoding a protein of the *ETT/ARF4* clade that lacked domains III and IV was measurably expressed in the ovary wall and ovules of all the angiosperm species analyzed, representing lineages that diverged at the base of angiosperm phylogeny. These results are consistent with a possible role for the truncation of ARF proteins in early carpel evolution.

The truncation of *ETT* is necessary for normal carpel development in *A. thaliana*.

The expression of truncated *ETT* or *ARF4* transcripts in the carpel tissues of widely diverged angiosperms led us to test the functional significance of such truncations for carpel development. Accordingly, we assessed the ability of various wild-type, chimeric and truncated versions of the *ETT* and *ARF4* coding sequences from *A. thaliana* to complement the *ett-1* mutation when expressed, downstream of the *A. thaliana ETT* promoter and 5'-leader sequence (*proETT*), in transgenic plants. Chimeric constructs were made by recombination of the *ETT* and *ARF4* coding sequences at a point of clear alignment between these, slightly upstream of domain III in *ARF4*, while truncated constructs were generated by the introduction of a stop codon at that same point (Fig 3A). As *ett-1* is infertile, we transformed a segregating population of *ett-1* mutants and screened the resulting T_1 transformants by PCR for homozygous *ett-1* mutants that were hemizygous for the inserted transgenes. Phenotypes were assessed in 10 to 20 such plants for each construction tested. In the case of T_1 plants that were partially or completely fertile, seed was collected and gynoecium phenotypes were also assessed in T_2 individuals testing positive for the transgene construction.

We found that a construction containing the wild-type *ETT* coding sequence (Fig 3A, *proETT::ETT*) completely restored wild-type gynoecium development to all 12 *ett-1* homozygotes analyzed in a T_1 population (Table 1, Fig 3B). These transformants were all highly fertile and gave T_2 seed which produced plants also showing wild-type gynoecium development. This positive control construction thus established that the cis-acting regulatory sequences used in these experiments were functional, and provided a standard against which the complementation achieved using other coding sequences could be measured.

Constructions containing either the wild-type *ARF4* coding sequence (Fig 3A, *proETT::ARF4*), or a translational fusion encoding the DNA binding domain of ETT together with domains III and IV of *ARF4* (Fig 3A, *proETT::ETT-ARF*), were able to partially restore the wild-type gynoecium phenotype to *ett-1* mutants (Table 1, Fig 3B). Gynoecia of the majority of T_1 transformants containing either of these constructions showed valves that were larger than those of *ett-1* mutants, but smaller that those of wild-type plants. The valves of these transformants were united by a replum of near wild-type size, which thus protruded significantly beyond the reduced valve tissues (Fig 3B). The stigmatic tissues of *proETT::ARF4* and *proETT::ETT-ARF4* transformants were intermediate in size between those of wild-type plants and *ett-1* mutants. These partially complemented transformants showed reduced fertility by comparison to wild-type, though did set some seed. The remaining T_1 transformants carrying either of these two constructions possessed gynoecia of near wild-type appearance, similar to those rescued using a *proETT::ETT* construction. However, in the T_2 generation, all transformants containing either the *pETT::ARF4* or *proETT::ETT-ARF4* constructions showed a partially complemented gynoecium phenotype with reduced value tissues, a protruding replum and increased stigmatic tissues, resembling that of the majority of T_1 transformants (Table 1).

The majority of T_1 plants transformed either with a chimeric construction encoding the DNA-binding domain of *ARF4*, translationally fused to the N-terminus of *ETT* (Fig 3A, *proETT::ARF4-ETT*), or with a truncated *ARF4* construction (Fig 3A, *proETT::ARF4-trunc.*), showed little if any restoration of wild-type gynoecium development (Table 1, Fig 3B). A minority of plants transformed with either of these two constructions produced some flowers containing gynoecia of near wild-type appearance, and which were able to set seed. In the T_2

generation, however, the progeny of these plants reverted almost completely to the *ett-1* phenotype.

The results of these analyses indicate a series of strength for the different constructions tested in their ability to complement the *ett-1* mutation. Accordingly, the chimeric *ETT-ARF4* and native *ARF4* coding sequences, both of which contained domains III and IV, showed similar capacities to complement *ett-1* mutants. However, the removal of sequences encoding domains III and IV showed opposite effects on the capacity of these two constructions to rescue *ett-1* mutants: the activity of ETT was increased by the removal of domains III and IV, whereas that of ARF4 was severely diminished. We conclude that the naturally occurring truncation of *ETT* is functionally significant for gynoecium development, possibly by rendering the ETT protein insensitive to negative regulation by Aux/IAA proteins. However, the removal of domains III and IV from ARF4 reduced the capacity of this transcription factor to functionally replace ETT when expressed from the *ETT* promoter, demonstrating the ETT and ARF4 DNA binding domains to be not entirely functionally equivalent.

Discussion

A reconstruction of the structural evolution of *ETT* and *ARF4* in the angiosperms.

The structural and phylogenetic analysis of *ETT* and *ARF4* orthologs from diverse seed plants has permitted us to reconstruct their molecular evolution (Fig 4), focusing particularly on truncation events that removed domains III and IV from the proteins they encode. According to our reconstruction (Fig 4), *ETT* and *ARF4* were derived by the duplication of a single *ARF3/4* gene in the angiosperm lineage, prior to the last common ancestor of the extant angiosperms. As non-truncated *ETT* and *ARF4* genes are present in species whose lineages diverged at the base of the angiosperm tree, neither *ETT* nor *ARF4* can, however, have been permanently truncated in the last common ancestor of the extant angiosperms. Following early speciation events in the flowering plant clade, which generated the Amborellales and Nymphaeales lineages, a truncation event occurred to *ETT* in the remaining common lineage of angiosperms by the formation of a stop codon, upstream of its tenth intron, thereby permanently removing domains III and IV from the ETT protein. This truncated *ETT* lineage has been inherited by both the monocots and eudicots and has become duplicated in the former group to generate, for example, four *ETT* paralogs in *O. sativa*. *ARF4*, by contrast, did not undergo a molecular truncation in the common lineage shared by the monocots and eudicots, and consequently encodes a full-length ARF protein in the eudicots. However, *ARF4*

appears to have been lost from the *O. sativa* lineage in the monocots, subsequent to the separation of the monocots and dicots. As no monocot *ARF4* orthologs have yet been reported, this gene loss may have occurred relatively early in monocot evolution, though the analysis of basal monocot lineages will be necessary to throw more light on this question. In the *C. aquatica* (Nymphaeales) lineage, *ARF4* has become permanently truncated by the generation of a stop codon, upstream of codon 11, while the *ETT* ortholog in this plant lineage has remained intact. The truncation of *ARF4* during the evolution of Nymphaeales can be assumed to have caused the inactivation of the domains III and IV that it previously encoded, leading to the subsequent erosion of sequence similarity to these conserved motifs. In *A. trichopoda*, the only living representative of Amborellales, an alternative splicing mechanism exists to generate an *ARF4* transcript that encodes a protein lacking domains III and IV, in addition to a low level of a full-length transcript. As in *C. aquatica*, an *ETT* transcript, encoding a protein containing domains III and IV, is also produced in this species.

A hypothesis for the role of ARF protein truncations in the origin of the angiosperms.

The importance of the truncation of ETT for carpel development in *A. thaliana* (Fig 3), combined with the presence of truncated *ETT* or *ARF4* transcripts in plant lineages that separated at a very early stage during angiosperm evolution (Fig 4), suggests a role for such truncated ARF proteins in the origin of the carpel. As *ETT* and *ARF4* were generated by a pre-angiosperm gene duplication, and have conserved a considerable degree of functional redundancy throughout angiosperm evolution (12), it is reasonable to suppose that the ETT and ARF4 proteins of early angiosperms were highly redundant. It follows that any effect produced by truncating one of these genes in the ancestor of the angiosperms could equally have been produced by truncating the other. However, we have shown that neither *ETT* nor *ARF4* can have been permanently truncated in the ancestor of the angiosperms (Fig 4). The only mechanism, of those uncovered in the present work, which could have generated ETT or ARF4 proteins that lacked domains III and IV in that ancestral species is the alternative splicing mechanism which still present in the most basal extant angiosperm lineage, Amborellales.

We propose that the removal of domains III and IV from the ARF4 protein occurred in an ancestor of the angiosperms by the alternative splicing of *ARF4* transcripts, following the gene duplication that separated the *ETT* and *ARF4* lineages. This alternative splicing mechanism may have evolved by nucleotide changes that generated Exonic Splicing Silencer

(ESS) sequences, associated with exon 11 of *ARF4*. ESS sequences are known from animal systems (27), and the general conservation of splicing phenomena between plants and animals (28) suggests that equivalent mechanisms exist in plants. As a consequence of the loss of exon 11, the shorter isoform of the resulting ARF4 protein would have been insensitive to regulation by Aux/IAA proteins- a condition which we postulate to have been necessary for the evolution of the carpel. Both ETT and the longer isoform generated of ARF4, by contrast, would have retained the capacity to be regulated by Aux/IAA proteins, which may have been necessary for the maintenance of their pre-existing functions. The production of either ARF4 or ETT proteins that lacked domains III and IV subsequently became fixed in distinct angiosperm lineages through two different mechanisms: one involving the permanent truncation of *ARF4* in Nymphaeales, and the other involving the permanent truncation of *ETT* in the common lineage leading to the monocots and eudicots (Fig 4). Thus, the respective genetic truncations of *ARF4* in the Nymphaeales, and of *ETT* in the monocots and eudicots, may represent parallel changes that resulted in the fixation a pre-existing molecular characteristic of importance to carpel development. Only in *A. trichopoda* has the ancestral mechanism of alternative splicing been conserved to generate ARF4 molecules that lack domains III and IV.

According to the above hypothesis, a vital genetic function for carpel development would have passed from *ARF4* to its paralog *ETT* in the common lineage leading to the monocots and eudicots, while *ARF4* orthologs would have retained their common ancestral function in carpel development in both the Amborellales and Nymphaeles lineages. Similar examples of the exchange of pre-existing functions between paralogous transcription factors have been reported. For example, the C-function MADS box genes *AGAMOUS* in *A. thaliana*, and *PLENA* in *Antirrhinum majus*, show very similar knockout mutant phenotypes, but are not orthologous. Instead, these genes are thought to be descended, in their respective plant lineages, from different members of a pair of paralogs that was generated by a duplication event in a common ancestor of the core eudicots.

Biochemical effects of the truncation of ARF proteins.

The *in planta* domain swapping experiments performed in the present work (Fig 3) have indicated the importance of the *ETT* truncation for gynoecium development in *A. thaliana*. However, a further effect was revealed in these experiments: the DNA-binding domains of ETT and ARF4 proved to be functionally equivalent *in planta*, conditionally on the presence

of domains III and IV. These data suggest a further subtlety to the evolution of ARF proteins by truncation: some ARF proteins may be inefficient as transcription factors if they lack the ability to dimerize via domains III and IV, whereas other may be able to function efficiently without such dimerization (3). It is possible, for example, that ARF proteins that have become permanently truncated over the course of evolution (eg ETT from *A. thaliana* and Caq-ARF4 from *C. aquatica*) have evolved to be capable of binding their target genes as monomers. Ulmasov *et al.* (10) have performed physical assays of DNA binding of ARF transcription factors with and without domains III and IV, and have found the presence of these domains to influence DNA binding capacity in some cases. Further studies will be required to test whether complementary changes to DNA binding domains may have occurred, following the evolutionary truncation of ETT and ARF4 proteins in diverse angiosperm groups.

Materials and Methods

Plant material.

Tissues of *A. trichopoda* were field-collected from Col d'Amieu (New Caledonia), and those of *E. distachya* were obtained from plants of the living collection of Lyon Botanic Garden (France). Plants of *C. aquatica* were obtained from Anthias S.A. (Les Chères, France) and grown in a freshwater aquarium. Seed of the Columbia-0 (Col-0) and Wassilewskija-2 (Ws-2) ecotypes of *A. thaliana* and of the *ett-1* mutant, in a Ws-2 background, were obtained from the Nottingham *Arabidopsis* Stock Centre (UK). *A. thaliana* plants, potted in peat-based compost, were grown to maturity in a growth chamber at 20°C under 16 h light/8 h dark cycles.

Gene identification and expression analyses

cDNAs corresponding to *ETT* and/or *ARF4* orthologs were obtained by screening previously described cDNA libraries constructed from *A. trichopoda* and *C. aquatica* flower RNAs (29), and a similarly constructed cDNA library from mixed developmental stages of male and female reproductive structures of *E. distachya.* A partial cDNA corresponding to *Atr-ARF4* was found by BLAST searching of an *A. trichopoda* flower EST database (http://pgn.cornell.edu/blast/blast_search.pl) and its sequence was then completed by RACE PCR from flower mRNA of *A. trichopoda.* Genomic loci were amplified by PCR using primers derived from the corresponding full-length cDNA sequences. Northern blot hybridizations, using radiolabeled probes, and *in situ* hybridizations, using digoxygenin-labeled probes, were performed as previously described (29).

Phylogenetic analyses.

Predicted amino acid sequences were aligned using MUSCLE (30) and manually adjusted. Selected blocks of sequences were then used in phylogenetic reconstructions. Maximum-likelihood and Bayesian inferences were performed using PhyML (31) and MrBayes (32), assuming in each case a WAG+I+Γ_4 model of nucleotide substitution. 500 bootstrap replicates were performed.

Production and analysis of transgenic plants.

Full-length *A. thaliana ETT* and *ARF4* coding sequences were amplified by RT-PCR from flower RNA of wild-type *A. thaliana* plants of the Col-0 ecotype. The resulting PCR products were then re-amplified using further primers to incorporate Gateway (Invitrogen) *attB1* and *attB2* recombination sites immediately upstream and downstream, respectively, of their initiation and stop codons. Chimeric coding sequences, recombined between the sequences encoding the DNA-binding and protein-protein (III and IV) interaction domains of *ETT* and *ARF4*, were produced by separately amplifying the 5'- and 3'-extremities of the *ETT* and *ARF4* full-length coding sequences using chimeric internal primers, and combining the resulting PCR products by further rounds of PCR amplification using the appropriate terminal primers. Truncated coding sequences were generated by re-amplifying cDNAs using *attB2* site-containing primers incorporating a stop mutation. The resulting wild-type, chimeric and truncated coding sequences were incorporated, using Gateway "BP" (Invitrogen) recombination reactions, into the *pDONR207* vector (Invitrogen) and their nucleotide sequences were fully determined. In parallel, a fragment carrying the presumed promoter and 5'-leader of *ETT (proETT)* was amplified from genomic DNA of the Col-0 ecotype of *A. thaliana*. One of the PCR primers used in this amplification incorporated an *Xba* I site immediately upstream of the ATG start codon of the main ORF in *ETT*, while the other straddled a naturally occurring *Xba* I site at -3736 bp, relative to the *ETT* start codon. The resulting promoter DNA was fully sequenced in a plasmid cloning vector, then released by treatment with *Xba* I and inserted into the unique *Xba* I site of a previously described plant transformation vector carrying a Gateway (Invitrogen) recombination cassette, upstream of a Nopaline Synthase terminator sequence (33). Wild-type, chimeric and truncated *ETT* and *ARF4* coding sequences from *pDONR207*-derived plasmids were then transferred to the resulting *proETT*-containing plant transformation vector by Gateway "LR" (Invitrogen) recombination reactions. This procedure generated a finished series of plant transformation

plasmids which were introduced into *Agrobacterium tumefaciens* strain *C58pmp90* by electroporation and then into a segregating population of *A. thaliana ett-1* mutants by the floral dip method (34). T_1 seed, harvested from dipped (T_0) plants, were selected on MS medium containing 8 mg/l of ammonium glufosinate (BASTA). The *ETT* locus of the BASTA-resistant T_1 plants selected was characterized by genomic PCR using combinations of the *ETT* coding sequence primers 5'-CTCGATGTTAAGCTTCACG and 5'-GCACTCCACCCGGTAGTGAGC, which flank the T-DNA insertion in the *ett-1* allele, and a primer derived from the inserted T-DNA sequence. The combined use of the above *ETT* primers enabled the discrimination, on the basis of the size of the PCR products generated, between the wild-type *ETT* allele and *ETT*-containing transgenes, as the latter lacked introns.

Acknowledgements

We thank the technical and secretarial staff of the RDP Laboratory for their general assistance and Lyon Botanic Garden for providing access to their living collection for sampling purposes. This project was financially supported by core funding from the RDP Laboratory's parent organizations: CNRS, INRA, ENS de Lyon and Université Claude-Bernard, Lyon-1.

References

1. Sessions, R. A. & Zambryski, P. C. (1995) Arabidopsis Gynoecium Structure in the Wild-Type and in Ettin Mutants. *Development* **121,** 1519-1532.
2. Sessions, A., Nemhauser, J. L., McColl, A., Roe, J. L., Feldmann, K. A. & Zambryski, P. C. (1997) ETTIN patterns the Arabidopsis floral meristem and reproductive organs. *Development* **124,** 4481-4491.
3. Guilfoyle, T. J. & Hagen, G. (2007) Auxin response factors. *Current Opinion in Plant Biology* **10,** 453-460.
4. Ulmasov, T., Hagen, G. & Guilfoyle, T. J. (1997) ARF1, a transcription factor that binds to auxin response elements. *Science* **276,** 1865-1868.
5. Reed, J. W. (2001) Roles and activities of Aux/IAA proteins in Arabidopsis. *Trends in Plant Science* **6,** 420-425.
6. Ulmasov, T., Hagen, G. & Guilfoyle, T. J. (1999) Activation and repression of transcription by auxin-response factors. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 5844-5849.
7. Tiwari, S. B., Wang, X. J., Hagen, G. & Guilfoyle, T. J. (2001) AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell* **13,** 2809-2822.
8. Dharmasiri, N., Dharmasiri, S. & Estelle, M. (2005) The F-box protein TIR1 is an auxin receptor. *Nature* **435,** 441-445.
9. Kepinski, S. & Leyser, O. (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature* **435,** 446-451.
10. Ulmasov, T., Hagen, G. & Guilfoyle, T. J. (1999) Dimerization and DNA binding of auxin response factors. *Plant Journal* **19,** 309-319.

11. Guilfoyle, T. J., Ulmasov, T. & Hagen, G. (1998) The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cellular and Molecular Life Sciences* **54,** 619-627.
12. Pekker, I., Alvarez, J. P. & Eshed, Y. (2005) Auxin response factors mediate Arabidopsis organ asymmetry via modulation of KANADI activity. *Plant Cell* **17,** 2899-2910.
13. Williams, L., Carles, C. C., Osmont, K. S. & Fletcher, J. C. (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and ARF4 genes. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 9703-9708.
14. Nishimura, T., Wada, T., Yamamoto, K. T. & Okada, K. (2005) The Arabidopsis STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. *Plant Cell* **17,** 2940-2953.
15. Nishimura, T., Wada, T. & Okada, K. (2004) A key factor of translation reinitiation, ribosomal protein L24, is involved in gynoecium development in Arabidopsis. *Biochemical Society Transactions* **32,** 611-613.
16. Savard, L., Li, P., Strauss, S. H., Chase, M. W., Michaud, M. & Bousquet, J. (1994) Chloroplast and Nuclear Gene-Sequences Indicate Late Pennsylvanian Time for the Last Common Ancestor of Extant Seed Plants. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 5163-5167.
17. Goremykin, V. V., Hansmann, S. & Martin, W. F. (1997) Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: Revised molecular estimates of two seed plant divergence times. *Plant Systematics and Evolution* **206,** 337-351.
18. Davies, T. J., Barraclough, T. G., Chase, M. W., Soltis, P. S., Soltis, D. E. & Savolainen, V. (2004) Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 1904-1909.
19. Kuzoff, R. K. & Gasser, C. S. (2000) Recent progress in reconstructing angiosperm phylogeny. *Trends in Plant Science* **5,** 330-336.
20. Aoki, S., Uehara, K., Imafuku, M., Hasebe, M. & Ito, M. (2004) Phylogeny and divergence of basal angiosperms inferred from APETALA3- and PISTILLATA-like MADS-box genes. *Journal of Plant Research* **117,** 229-244.
21. Stellari, G. M., Jaramillo, M. A. & Kramer, E. M. (2004) Evolution of the APETALA3 and PISTILLATA lineages of MADS-box-containing genes in the basal angiosperms. *Molecular Biology and Evolution* **21,** 506-519.
22. Zanis, M. J., Soltis, D. E., Soltis, P. S., Mathews, S. & Donoghue, M. J. (2002) The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 6848-6853.
23. Qiu, Y. L., Lee, J., Whitlock, B. A., Bernasconi-Quadroni, F. & Dombrovska, O. (2001) Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Molecular Biology and Evolution* **18,** 1745-1753.
24. Moore, M. J., Bell, C. D., Soltis, P. S. & Soltis, D. E. (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 19363-19368.
25. Saarela, J. M., Rai, H. S., Doyle, J. A., Endress, P. K., Mathews, S., Marchant, A. D., Briggs, B. G. & Graham, S. W. (2007) Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* **446,** 312-315.
26. Sokoloff, D. D., Retnizowa, M. V., Macfarlane, T. D. & Rudall, P. J. (2008) Classification of the early-divergent angiosperm family Hydatellaceae: one genus

instead of two, four new species and sexual dimorphism in dioecious taxa. *Taxon* **57,** 179-200.
27. Wang, Z. F., Xiao, X. S., Van Nostrand, E. & Burge, C. B. (2006) General and specific functions of exonic splicing silencers in splicing control. *Molecular Cell* **23,** 61-70.
28. Irimia, M., Rukov, J. L., Penny, D. & Roy, S. W. (2007) Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *Bmc Evolutionary Biology* **7**.
29. Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C. & Scutt, C. P. (2005) Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 4649-4654.
30. Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* **5,** 1-19.
31. Guindon, S. & Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52,** 696-704.
32. Ronquist, F. & Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572-1574.
33. Fourquin, C., Vinauger-Douard, M., Chambrier, P., Berne-Dedieu, A. & Scutt, C. P. (2007) Functional conservation between CRABS CLAW orthologues from widely diverged Angiosperms. *Annals of Botany* **100,** 651-657.
34. Clough, S. J. & Bent, A. F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal* **16,** 735-743.

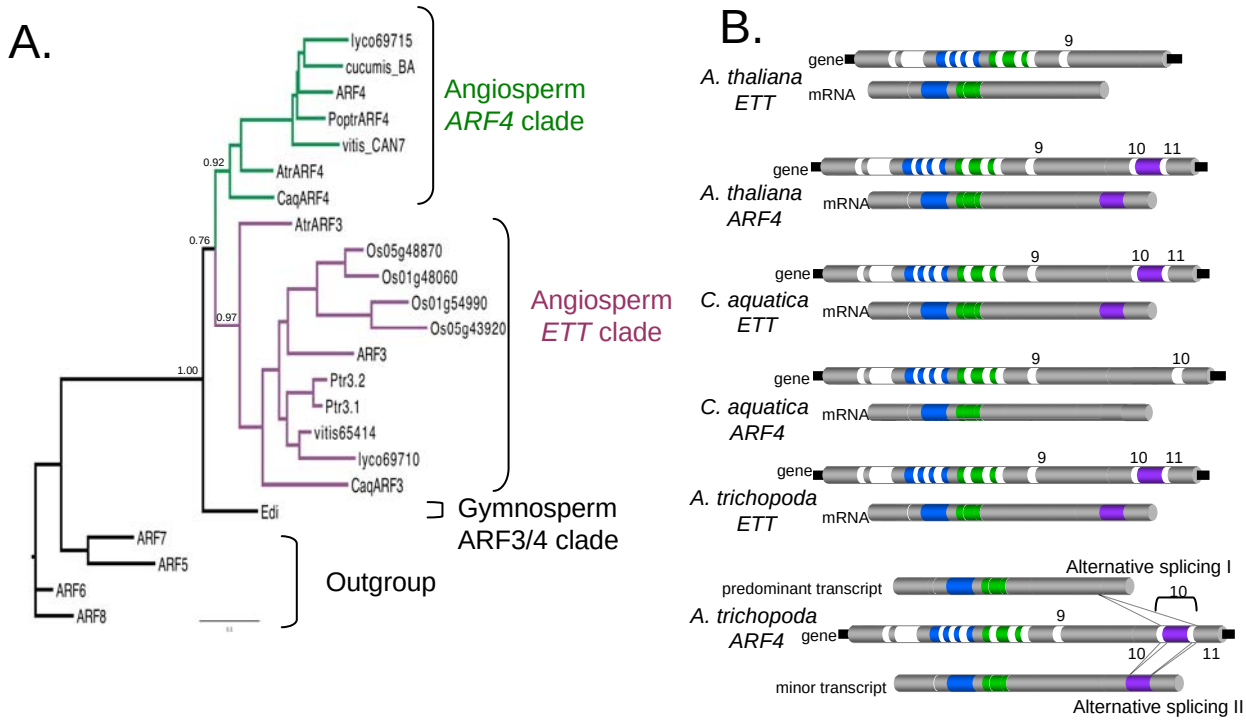
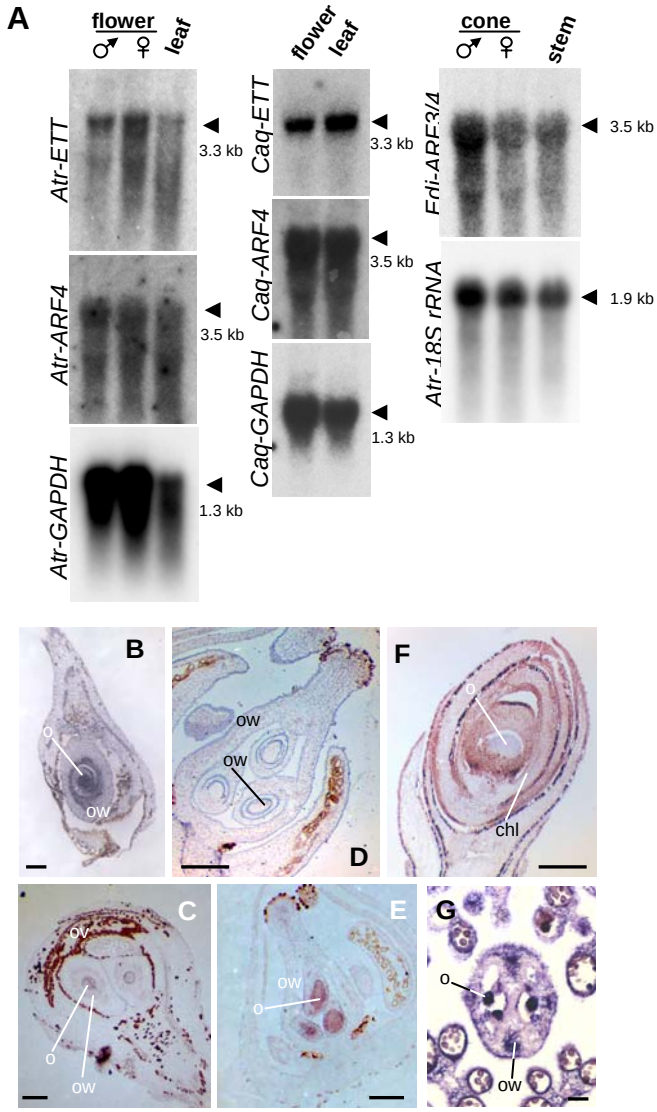


Figure 1 Phylogeny and gene structure of *ETT* and *ARF4* orthologs in the seed plants.

A. The ML phylogeny of the *ETT/ARF4* clade in the seed plants. *ETT* and *ARF4* from *A. thaliana* group into separate sister clades, which both contain sequences, in basal positions, from ANA grade angiosperms. A gymnosperm gene, *Edi-ARF3/4*, occurs in a sister position to the combined angiosperm *ETT* and *ARF4* clades. Accession numbers for the sequences included in this analysis are as follows: *lyco69715* (TC188486), *cucumis_BA* (BAD19065), *ARF4* (NP_200853), *PoptrARF4* (fgenesh4_pg.C_LG_IX001438), *vitis_CAN7* (CAO21817), *Os05g48870* (AB071290), *Os01g48060* (NM_001050360), *Os01g54990* (NM_001050809), *Os05g43920* (AK067927), *ARF3* (NP_180942), *Ptr3.2* (fgenesh4_pg.C_scaffold_187000006), *Ptr3.1* (estExt_Genewise1_v1.C_LG_IV2935), *vitis65414* (CAN65414), *lyco69710* (TC175247), *ARF7* (NP_851047), *ARF5* (NP_173414), *ARF6* (NP_174323), *ARF8* (NP_198518). Bootstrap values are indicated at key nodes.

B. The structure of *ETT* and *ARF4* genes and their transcripts from *A. thaliana* and the ANA grade basal angiosperms *A. trichopoda* and *C. aquatica*. Introns are shown in white, sequences encoding the B3 and ARF regions of the DNA-binding domain are shown in green and blue, respectively, and sequences encoding domains III and IV, involved in protein-protein interactions, are shown mauve. Introns are numbered from Intron 9, whose position is conserved in all genes. Independent truncations that removed sequences encoding domains III and IV are apparent in *A. thaliana ETT* (which is representative of all eudicot and monocot *ETT* orthologs) and in *C. aquatica ARF4*. Two alternatively spliced mRNA forms are shown for *Ath-ARF4* from *A. trichopoda*, the shorter of which results from the retention of Exon 11 between two introns.


A
flower
♂
♀
leaf
Atr-ETT
3.3 kb
Atr-ARF4
3.5 kb
Atr-GAPDH
1.3 kb
flower
leaf
Caq-ETT
3.3 kb
Caq-ARF4
3.5 kb
Caq-GAPDH
1.3 kb
cone
♂
♀
stem
Fdi-ARF3/4
3.5 kb
Atr-18S rRNA
1.9 kb
B
o
ow
ow
D
F
o
chl
C
ow
o
ow
E
ow
o
G
o
ow

H

Transcript	*ta-siR2141* site	*ta-siR2142* site
Ath-ETT	AGGGUCUUGCAAGGUCAAGAA	AAGGUCUUGCAAGGUCAAGAA
Caq-ETT	AAGGUCUUGCAAGGUCAAGAA	AAGGUCUUGCAAGGUCAAGAA
Atr-ETT	AAGGUCUUGCAAGGUCAAGAA	AAGGUCUUGCAAGGUCAAGAA
Ath-ARF4	AAGGUCUUGCAAGGUCAAGAA	AGGGUCUUGCAAGGUCAAGAA
Caq-ARF4	AAGGUCUUGCAAGGUCAAGAA	AAGGUCUU**U**CAAGGUCAAGAA
Atr-ARF4	AAGGUCUUGCAAGGUCAAGAA	AAGGUCUUGCAAGGUCAAGAA
Edi-ARF3/4	AAGGUCUUGCAAGGUCAAGAA	AAGGUCUUGCAAGGUCAAGAA
	**********	******** **********
Consensus	**AAGGUCUUGCAAGGUCAAGAA**	**AGGGUCUUGCAAGGUCAAGAA**

I

Transcript	Length of 5’-leader analyzed	uORFs identified
Ath-ETT	365 bp	2
Caq-ETT	194 bp	7
Ath-ARF4	511 bp	4
Atr-ARF4	115 bp	1
Edi-ARF3/4	390 bp	6

Figure 2. mRNA expression patterns of *ETT/ARF4* clade genes in the seed plants and motifs associated with their post-transcriptional and translational regulation.

A. Northern blot hybridizations showing expression of *ETT/ARF3* clade genes in reproductive and vegetative tissues of *A. trichopoda (Atr-ETT* and *Atr-ARF4)*, *C. aquatica (Caq-ETT* and *Caq-ARF4)* and *E. distachya (Edi-ARF3/4).* Hybridizations of nominally constitutive *GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE* (*GAPDH)* or 18S ribosomal RNA probes are included to show approximately equal loading of tracks. The molecular sizes of the hybridizing bands are shown.

B to G. Non-radioactive *in situ* hybridizations to female reproductive tissues showing: expression of (B.) *Atr-ETT* and (C.) *Atr-ARF4* in the ovule and ovary wall of *A. trichopoda*; (D.) uniformly low expression of *Caq-ETT* in carpel tissues of *C. aquatica*; (E.) expression of *Caq-ARF4* in the ovules and ovary wall of *C. aquatica; and* (F.) expression of *Edi-ARF3/4* in the ovule and chlamys of *E. distachya.* The expression of *ETT* (G.) in ovules and the ovary wall of *A. thaliana,* is shown for comparison. (o = ovule, ow = ovary wall, chl = chlamys. Scale bars represent 500μm).

H. Conservation of putative sites of regulation by ta-siRNAs in *ETT/ARF4* orthologs from *A. thaliana, A. trichopoda, C. aquatica* and *E. distachya.*

I. Numbers of upstream ORFs (uORFs) identified in the 5’-leader sequences of *ETT/ARF4* orthologs from *A. thaliana, A. trichopoda, C. aquatica* and *E. distachya.*

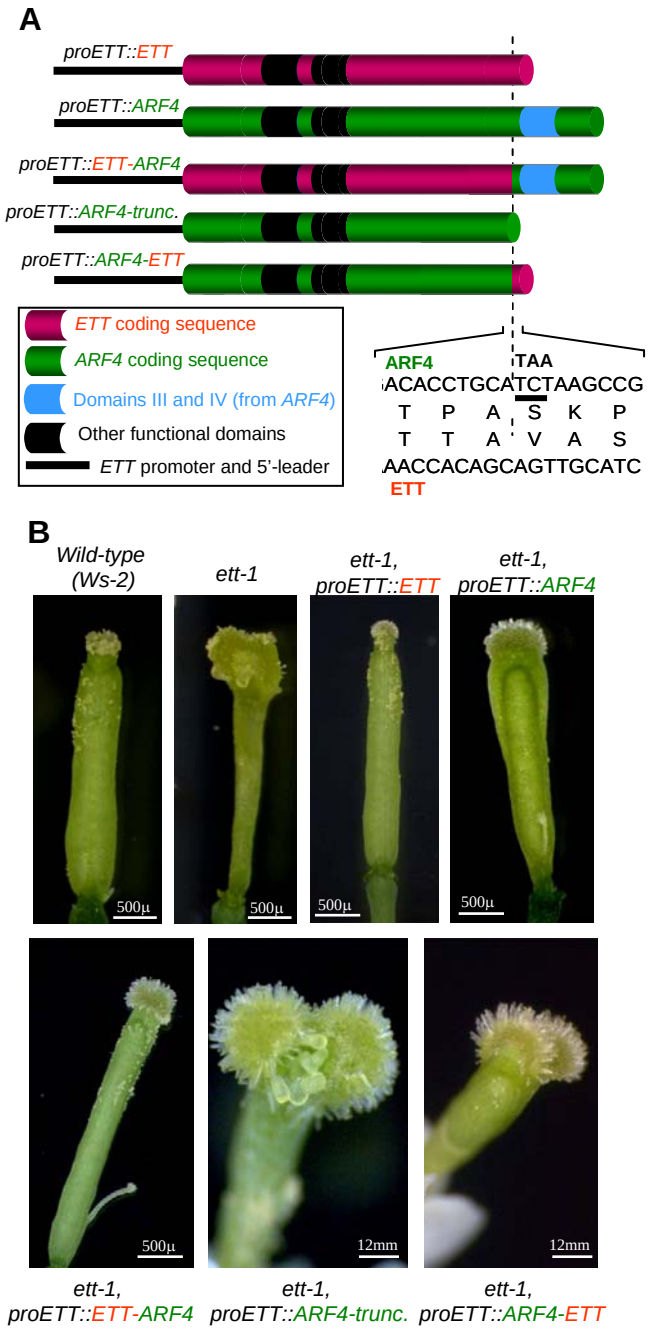


Figure 3. The effect of ARF protein truncations on gynoecium development in *A. thaliana.*

A. Transgene constructions containing wild-type, chimeric and truncated *ETT* and *ARF4* coding sequences, inserted downstream of the *ETT* promoter and 5'-leader sequence. The sequences surrounding points of recombination (dashed line) and truncation, by the insertion of a TAA stop codon, are shown.

B. Gynoecia of: wild-type *A. thaliana*, the *ett-1* mutant, and typical *ett-1* mutants transformed with the constructions shown in Fig 3A.

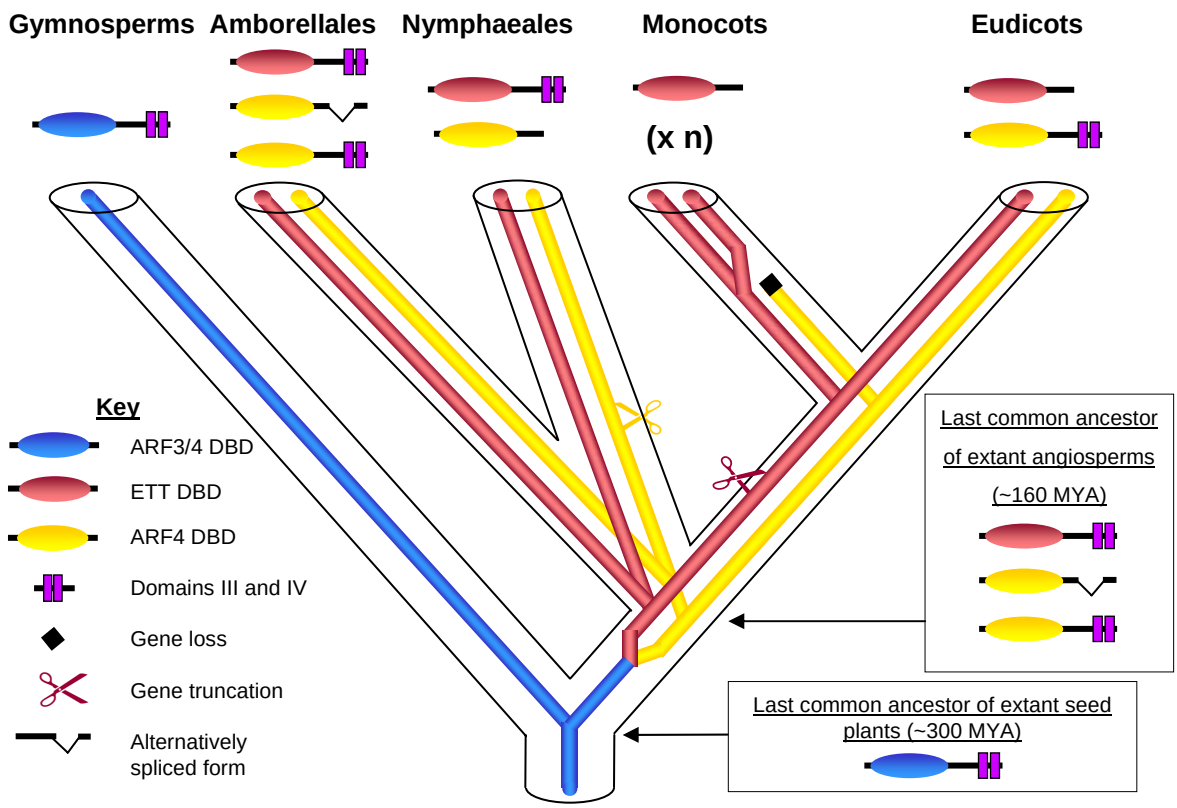


Figure 4. A model for the structural evolution of ETT and ARF4 proteins in the angiosperms, incorporating a hypothesis for the role of the alternative splicing of *ARF4* in the origin of the carpel. According to this model, the production an ARF4 protein isoform that lacked domains III and IV was present in the last common ancestor of the extant angiosperms. The production of truncated ARF molecules, necessary for carpel development, subsequently became fixed by distinct genetic truncation events affecting *ARF4* in the Nymphaeales lineage and *ETT* in a common lineage shared by the monocots and eudicots. The ancestral mechanism of alternative splicing has been maintained in the earliest-diverging angiosperm lineage, Amborellales.DBD = DNA Binding Domain.

Transgene constructions	Numbers of initial transgenic lines, fully, partially, or not complemented						
	T_1 Generation			T_2 Generation			
	full	partial	none	full	partial	none	Not tested (sterile in T1)
proETT::ETT	**12 (100%)**	0 (0%)	0 (0%)	**12 (100%)**	0 (0%)	0 (0%)	0 (0%)
proETT::ARF4	1 (10%)	**9 (90%)**	0 (0%)	0 (0%)	**10 (100%)**	0 (0%)	0 (0%)
proETT::ETT-ARF4	5 (26%)	**14 (74%)**	0 (0%)	1 (5%)	**18 (95%)**	0 (0%)	0 (0%)
proETT::ARF4-trunc.	5 (038%)	0 (0%)	**8 (62%)**	1 (8%)	0 (0%)	**4 (31%)**	8 (62%)
proETT::ARF4-ETT	5 (33%)	0 (0%)	**10 (67%)**	1 (6%)	0 (0%)	**4 (27%)**	10 (67%)

Table 1. Complementation of homozygous *ett-1* mutant plants using transgene constructions containing wild-type, chimeric and truncated versions of the *ETT* and *ARF4* coding sequences. The most frequent phenotype in each generation is indicated in bold.

III) Perspectives

Le travail présenté ci-dessus semble indiquer une importance évolutive de la troncation et de la perte des domaines III et IV chez un des gènes du clade ARF3-4. Dans le contexte des connaissances actuelles, la perte des domaines III et IV se traduit par une incapacité de former des dimères avec les autres ARFs et les AUX/IAAs. Or, il est connu que les ARFs reconnaissent et se fixent sur les séquences AuxREs (dans les promoteurs des gènes de réponse précoce à l'auxine) sous la forme de dimères (Ulmasov et al., 1999) d'ARFs interagissant via leurs domaines III et IV. Nous pouvons donc émettre l'hypothèse qu'un facteur de transcription ARF tronqué doit réguler ses gènes cibles de manière moins efficace. Il serait intéressant de quantifier l'efficacité de cette régulation transcriptionnelle des constructions ARF4-trunc, ARF4-ETT et ETT-ARF4 chez *A. thaliana*. On pourrait imaginer de synthétiser ces protéines *in vitro* et de tester leur interaction sur des séquences AuxREs par des expériences de gel retard.

Il serait intéressant de cloner les orthologues des gènes *ARF3* et *ARF4* chez un plus grand nombre d'angiospermes dont les lignages sont apparus assez tôt au cours de l'évolution. Je pense notamment à des espèces appartenant aux Austrobaileyales (grade ANA), Eumagnoliids et Ranunculales. Concernant les Austrobaileyales, j'ai commencé à travailler chez *Illicium parviflorum* chez qui j'ai cloné un orthologue du gène *ARF3*, possédant les domaines III et IV comme il semble être le cas dans l'ensemble du grade ANA. Malheureusement, je n'ai cloné que la partie 5' d'un gène qui est apparenté au lignage *ARF4* (selon les premiers hits BLAST) et qui n'apporte pas de renseignements sur la présence/absence des domaines III et IV en 3' de la molécule.

Chapitre 3 : GÈNES *ARF* ET ÉVOLUTION DES EMBRYOPHYTES

I) Introduction et résumé de l'article

Cet article résume un travail qui se veut à la fois dans la continuité des travaux sur l'évolution moléculaire des gènes *ARF3* et *ARF4*, mais aussi dans la généralisation du rôle déterminant des gènes *ARFs* dans l'évolution des embryophytes. Chez les angiospermes, le clade ARF3/ARF4 regroupe les gènes *ARF3* et *ARF4* dont l'un des deux code pour une protéine tronquée (sans domaines III et IV). Chez les angiospermes basales, la troncation affecte le lignage codant pour ARF4 alors que c'est la protéine ARF3 qui est plus courte chez les monocots et les eudicots supérieures. Cette convergence évolutive suggère que la présence d'une protéine ARF3 ou ARF4 tronquée constitue une véritable contrainte développementale

ou une innovation évolutive. De manière encore plus intéressante, les mécanismes qui ont conduit à ces formes protéiques tronquées sont différents. Il peut s'agir d'une véritable troncation au niveau génomique (*Cabomba aquatica*) ou bien d'un épissage alternatif qui supprime les domaines III et IV au niveau du transcrit mature (*Amborella trichopoda*).

Au regard de ces résultats, la question qui s'est naturellement posée était de savoir si la troncation jouait un rôle plus général au sein de la famille multigénique ARF (en dehors du clade ARF3/ARF4) et ceci chez l'ensemble des bryophytes. D'autant plus que les génomes de la bryophyte *P. patens* et de la lycophyte *S. moellendorffii* venaient d'être déposés dans les bases de données publiques. Cependant, peu ou pas de séquences ARF étaient connues dans le grade ANA ou des embranchements entiers tels que les gymnospermes.

Cette analyse globale a nécessité : (i) le clonage de gènes *ARF* chez des espèces du grade ANA et des gymnospermes, (ii) le séquençage intégral de clones (obtenus sur demande) ayant servi à la synthèse d'ESTs, (iii) la recherche *in silico* de gènes ARF dans des génomes entiers ou des bases d'ESTs. Dans un deuxième temps, j'ai reconstruit la phylogénie de la famille ARF au sein des embryophytes, ce qui m'a permis de placer les événements évolutifs majeurs comme les duplications. Ce travail a par ailleurs été très utile pour obtenir une nomenclature (ou classification) des gènes *ARF* basée sur des critères phylogénétiques et non arbitraires. En effet, les gènes *ARF* ont été nommés de 1 à 23 chez *A. thaliana* selon l'ordre dans lequel ils ont été identifiés.

Initialement, au moins trois gènes *ARF* étaient présents chez l'ancêtre commun des embryophytes actuelles. Cette famille a par la suite connu plusieurs vagues de duplication au sein des angiospermes d'où un nombre élevé de gènes *ARF* chez *O. sativa* (25 membres), *P. trichocarpa* (39 membres), et *A. thaliana* (23 membres). On détecte la présence de transcrits tronqués dont la répartition n'est ni clade spécifique ni espèce spécifique, suggérant des troncations indépendantes au cours de l'évolution. Sachant que les domaines III et IV sont connus pour intervenir dans la réponse à l'auxine, on peut se demander si la perte de ces domaines ne constitue pas une solution évolutive convergente d'échappement à cette voie de signalisation.

J'ai mené intégralement les différents axes de ce projet : la conception, le clonage des séquences ARFs chez les espèces du grade ANA et les gymnospermes, la reconstruction phylogénétique, l'analyse d'évolution moléculaire ainsi que la rédaction de ce manuscrit en anglais.

II) Article

Classification

Biological Sciences (Evolution)

Title

Phylogenomics of ARFs reflect the evolution of body plans in the land plants.

Author affiliation

Cédric Finet[1], Ferdinand Marlétaz[2], Annick Berne-Dedieu[1] and Charles P. Scutt[1]

[1] Université de Lyon, Laboratoire de Reproduction et Développement des Plantes, UMR 5667, CNRS, INRA, Université Claude Bernard Lyon I, IFR 128, Ecole Normale Supérieure de Lyon, 46 allée d'Italie 69364 Lyon Cedex 07, France

[2] Station Marine d'Endoume, UMR 6540, DIMAR, CNRS, Université de la Méditerranée, Chemin de la Batterie des Lions 13007 Marseille, France

The plant hormone auxin (indole-3-acetic acid) regulates many physiological and developmental processes of land plants, such as organogenesis, vascular tissue differentiation, cell elongation, apical dominance, gravitropism, and embryo and root patterning (reviewed by (1)). Particularly, when applied at early stages, both auxin and chemical inhibitors of auxin transport severely affect the architecture of the embryo (2, 3). As the characteristic body plans of the different phyla of land plants are established during embryonic and early postembryonic development, auxin may act as a critical developmental signal in this process. Changes in auxin perception and signaling could therefore have underpinned the diversification of body plans during the evolution of land plants (4).

To address this question, we have attempted to describe the diversity of the Auxin Response Factor (ARF) genes, that play a major role in auxin signaling, particularly in the context of land plant evolution. ARFs are transcription factors that trigger the expression of early auxin-responsive genes such as members of the *Aux/IAA*, *GH3* and *SAUR* gene families (5). ARFs act at the interface between a hormonal pathway, by being sensitive to auxin, and a transcriptional pathway, by binding to auxin-response elements (AuxREs) in the promoter of their target genes. *Arabidopsis thaliana* contains 23 *ARF* genes, *ARF1* to *ARF23* (6). Most of these encode proteins consisting of an amino-terminal B3-like DNA-binding region that includes an ARF family-specific domain, a variable middle region that confers activator or repressor activity, and a carboxy-terminal dimerization region (domains III and IV) involved in homo- and hetero-dimerization with other ARFs (7) (8).

At low auxin concentrations, ARFs are inhibited by interacting with proteins of the Aux/IAA family via domains III and IV, which are conserved between the two protein families (9). Aux/IAAs are small, short-lived nuclear proteins encoded by genes that were originally identified for their rapid upregulation in response to auxin (5). High auxin concentrations promote the degradation of Aux/IAA proteins, which can be hypothesized to release interacting ARFs from inhibition. Despite the fact that *Aux/IAA* genes could have been played a key role in the regulation of *ARFs*, their evolutionary origin is still unknown.

Our understanding of auxin signaling has been hampered by experimental difficulties because of both the large size of gene families involved in this process, and of the functional redundancy within these. We reasoned that a better understanding of ARF gene evolution would facilitate studies of auxin signaling. In the present work, we establish that coding sequence changes have made a substantial contribution to ARF evolution. We discuss this finding in the context of the debate over whether coding and regulatory mutations make qualitatively distinct contributions to phenotypic evolution.

Results

A first phylogeny of the ARF family across the land plants.

Here, we report the cloning of ARF sequences from two gymnosperms (*Ephedra distachya* and *Ginkgo biloba*) and three basal angiosperms belonging to the ANA grade (*Amborella trichopoda*, *Cabomba aquatica* and *Illicium parviflorum*) and present an analysis of a dataset of ARF proteins from a range of species that includes all major land plant groups (SI Table 1). Maximum-likelihood reconstructions of ARF phylogenies were congruent, and gave rise to trees very similar in topology to the Bayesian phylogeny in Fig. 1. The results of our phylogenetic analysis help to clarify the number of main clades in the ARF family. Accordingly, we propose the following nomenclature : clade A groups the ARF5, ARF7, ARF6, and ARF8 lineages ; clade B groups the ARF2, ARF1, ARF9, ARF3 and ARF4 lineages ; and clade C groups the ARF10, ARF16 and ARF17 lineages.

Inferring the minimum ARF complement of the last common ancestor of the extant land plants.

The original reports of recent plant genome projects include accounts of the ARF genes found. While data originally reported for *Populus trichocarpa* were accurate (10), the results obtained for the moss genome were not.

We have repeated the search for each available plant genome sequence, including those for which estimates had previously been made. This yields consistent results with those reported for *Arabidopsis thaliana* (23 ARF genes, (6)), *Oryza sativa* (25 ARF genes, (12)), and *Populus trichocarpa* (39 ARF genes, (10)). We present the first analysis of ARF genes for two recently sequenced species, the moss *Physcomitrella patens* (13 ARF genes, (11)) and the spikemoss *Selaginella moellendorffii* (10 ARF genes).

According to the phylogeny of the ARF family, there were at least 10 ARF genes in the last common ancestor of extant angiosperms. The three major clades we have identified, A, B and C, are present in both bryophytes and lycophytes, suggesting that these were present in the last common ancestor of the extant embryophytes. Therefore, we can assume that there were at least 3 ARF genes in the last common ancestor of extant embryophytes and vascular plants.

Preangiosperm gene duplication events within several ARF subfamilies.

In flowering plants, the clades ARF6 and ARF8 are phylogenetically related and both contain sequences from ANA grade angiosperms, monocots, and eudicots, suggesting that the first gene duplication event within the ARF6/8 subfamily occurred before the origin of the extant

angiosperms. This hypothesis is supported by the phylogenetic position of gymnosperm orthologs (*E. distachya ARF6/8*, *G. biloba ARF6/8*, *P. taeda ARF6/8.1* and *ARF6/8.2*) that occur in a sister position to the combined ARF6/8 clade (Fig 1).

Both *arf6* and *arf8* single mutants show delayed flower maturation and subtly reduced fertility, while flower development in *arf6/arf8* double mutants is arrested before maturity resulting in complete sterility (13). In addition to its redundant role with ARF6 in flower development, ARF8 plays an essential role in coupling fertilization and fruit development (14). Although ARF6 and ARF8 act largely redundantly, these two genes may have evolved some new functions by neofunctionalization.

Similarly, the clades ARF3 and ARF4 result from a preangiosperm gene duplication event. In *Arabidopsis*, single loss-of-function *arf3* (*ettin*) mutants show severe defects in the establisment of the polarity of the gynoecium (15), whereas no mutant phenotype has been observed in single *arf4* mutants. The *ett* phenotype is enhanced in *arf3 arf4* double mutants which exhibit reduced abaxial identity in all lateral organs, including leaves (16). These results indicate that the paralogs ARF3 and ARF4 are involved in the same developmental pathway.

Last, ARF5/MONOPTEROS (MP) is required for embryonic root initiation (17), whereas both MP and ARF7/NONPHOTOTROPIC HYPOCOTYL 4 (NPH4) contribute to cotyledon development (18).

Additional duplication events within the monocot and eudicots lineages.

In flowering plants, ARF families are typically large, with 23 members reported from the *A. thaliana* genome (6), and 25 and 39 from those of *O. sativa* (12) and *P. trichocarpa* (10), respectively. This phenomenon could be explained by widespread genome duplications throughout the history of angiosperms (19).

The origin of several cases of the post-transcriptional of ARF genes can be traced to lower embryophytes.

We have compiled a list showing the presence/absence of regulatory sites previously reported in the literature (SI Table 2). The distribution of the different forms of regulation is discussed in a phylogenetic framework. Our current view of the evolution of *ARF* regulation is summarized in Fig 2.

The *ARF2*, *ARF3* and *ARF4* Arabidopsis genes have been shown to be targeted by an endogenous trans-acting short-interfering RNA (*tasiR-ARF*) (20). Similarly to the miRNAs

pathway, *tasiR-ARF* is able to trigger the cleavage of *ARF2*, *ARF3/ETT*, and *ARF4* transcripts (20). According to the conservation of *tasiR-ARF* and its target genes in rice and maize, this postranscriptional regulation has been hypothesized to have originated before the separation of the monocot and dicot lineages (20). We identified *tasiR-ARF* target sites in orthologues of *ARF2*, *ARF3* and *ARF4* genes in the major groups of angiosperms, including the ANA grade, the monocot and dicot lineages. In addition, *tasiR-ARF* target sites are also found in *ARF3/4* and *ARF2/1/9* genes in several gymnosperm species, whereas no conserved sites have been identified in B clade homologues in mosses and spikemosses. These findings suggest that the regulation via *tasiR-ARF* arose in the lineage leading to the seed plants. Strikingly, tasiR-ARF target sites are absent from the ARF1/9 clade in angiosperms. Given that ARF1/9 and ARF2 are phylogenetically sister groups, *tasiR-ARF* target sites have probably been lost in ARF1/9 clade in the flowering plants.

In *Arabidopsis*, *ARF6* and *ARF8* are targets for *miR167*, whereas *ARF10*, *ARF16* and *ARF17* are targets for *miR160* (21). *miR160* target sites have been identified in all ARF genes of the C-clade and seem broadly conserved among the major phyla of embryophytes. Moreover, *MIR160* genes are present in *Physcomitrella* and *Selaginella* genomes, corroborating the hypothesis that C clade genes were regulated by *miR160* in the ancestor of the extant land plants. On the contrary, *miR167* target sites seem to be restricted to the seed plants. No *miR167* target sites have been revealed in the sister clade ARF5/7, nor in homologues in basal embryophytes, indicating that the regulation by *miR167* may have specifically appeared in the ARF6/8 lineage.

Upstream open reading frames (uORFs) are small open reading frames in the 5'-leader sequence of a mature mRNA, which can mediate translational regulation of the major ORF (mORF). If uORFs are recognized by a ribosome scanning the mRNA, translation will be terminated at the stop codon of the uORF, and translation of the downstream ORF will require the reinitiation of translation. The majority of uORFs appear to act in an amino acid sequence-independent manner. However, some uORFs do rely on peptide sequences to mediate translational regulation of the associated mORF. In *A. thaliana*, some uORFs have been predicted in the 5'-leader sequence of numerous ARF transcripts, and the occurrence of transcripts from the uORFs was experimentally confirmed (22). In particular, *ARF3* and *ARF5* have uORFs in their 5'-leader sequence that negatively regulate the translation of the mORF (23). Some uORFs have been identified in most of the ARF clades, which underlines that this regulation could be relevant at a macroevolutionary scale. Unfortunately, it remains difficult to reach a conclusion on the occurrence of uORFs in the C clade in *Arabidopsis ARF10*,

ARF16 and *ARF17*, due to the fact that the start of transcription has not yet been determined. This limitation is also true for the *Physcomitrella* and *Selaginella* genomes.

We investigated whether uORFs in the 5'-leader sequence of the orthologues were conserved at the amino acid level. Using this criterion, no conserved uORFs were identified in *O. sativa*, *P. trichocarpa* or *V. vinifera* genomes. These results are similar to those obtained by an exhaustive scan for conserved uORFs in angiosperms (24).

The regulation by uORFs, in an amino-acid sequence-independent manner, is higly frequent in the ARF family, suggesting that posttranscriptional regulation may be important for ARF gene expression.

Identification of a novel conserved domain in ARFs.

ARF proteins contain a variable domain located between DNA-binding domain and domain III, which has been hypothezised to confer activatior or repressor activity of a given ARF. Despite its relatively variable sequence, the middle domain of ARFs contains numerous conserved motifs. For example, recent studies revealed the presence of target sites for *miR160* in *ARF10, ARF16 and ARF17*, for *miR167* in *ARF6 and ARF8*, and for ta-siRNAs in *ARF2*, *ARF3* and *ARF4* (20). Here, we report the existence of a previously uncharacterized domain in the middle region of ARF proteins. This motif is located before the domain III in all ARF clades, except for ARF5 (SI Table 2). This domain is predicted to form a β-sheet secondary structure.

Independent truncations occurred during ARF family evolution.

In *A. thaliana*, four members of the ARF family - ARF3, ARF13, ARF17 and ARF23 - are truncated at the genomic level, leading to proteins lacking domains III and IV. In spite of these truncations, it has been shown that these proteins are functional (except for ARF23, which appears to be a pseudogene, (25)). A striking feature in the occurrence of this phenomenon is that truncations occurred independently both in diverse taxa and ARF clades (Fig. 3). For example, two sporadic occurrences of truncation event are apparent in clade C : one in a gene from *P. patens* (*P. patens ARF10/16/17.2*) and the other one in *ARF17* orthologues in flowering plants.

Alternative splicing is frequent in ARFs.

The establishment of a large ARFs dataset presented the opportunity to identify numerous examples of alternative transcripts in land plants. This identification required several steps :

first, we built clusters that gathered all transcripts from a single gene ; second, we searched for canonical splice sites recognition GT/AG at the ends of the indels ; third, assuming the conservation of intron positions between orthologues, we checked whether the indel corresponded precisely to an intron/exon junction in *A. thaliana*.

This method allowed us to reveal the occurrence of alternative splicing in different ARF lineages and different species (Fig 3). Interestingly, alternative splicing can act on different exons along the transcript. For instance, this mechanism can generate some transcription factors without the domains III and IV involved in dimerization (*A. trichopoda* ARF4), without the repressor domain (*S. lycopersicum* ARF2), without a serine-rich region in the N-terminal end (*A. thaliana ARF4*), or proteins for which the ARF domain is not entire (*G. biloba* ARF6/8, *A. thaliana* ARF13).

An example of distinct functions for two isoforms of the *ARF4* gene in *A. thaliana*.

The method we used to identify some alternative transcripts is not quantitative, and does not permit us to exclude aberrant or incompletely spliced mRNAs. Here, we tested the biological relevance of the existence of two isoforms of the *ARF4* mRNA, *ARF4db* and *isoformΔ(99-170)ARF4*. First, the *isoformΔ(99-170)ARF4* variant represents one-third of total *ARF4* transcripts in *Arabidopsis* (data not shown), challenging the idea that the *isoformΔ(99-170)ARF4* variant could be an aberrant transcript. Second, this alternative splicing leads to the loss of an S-rich region which is widely conserved between ARF4 orthologues, suggesting that this deletion may be of functional significance. We carried out an experiment to indirectly test the function of two isoforms of the *ARF4* mRNA, *ARF4db* (A line) and *isoformΔ(99-170)ARF4* (MA line). Indeed, *arf4* mutants do not exhibit visible phenotype in *Arabidopsis*, making not feasible direct complementation assays of *arf4* mutants. Consequently, we took advantage of the fact that a *pETT::ARF4db* transgene was found to partially restore the wild-type phenotype in 100 % of transformed *ett-1* mutants in a T_2 population of 40 plants. Carpels and siliques of these transformants were completely closed, but though contained enlarged replum and style tissues. Moreover, no significant variability was detected between *ett-1* and A lines in sepal number (Wilcoxon test, p-value = 0.4), or petal number (Wilcoxon test, p-value = 0.7), though stamen number proved more variable (Wilcoxon test, p-value = 1.2 e-6).

By contrast, a *pETT ::isoformΔ(99-170)ARF4* transgene dramatically enhanced the *ett-1* mutant phenotype (Fig 4). Indeed, defects in carpel polarity and fusion were accompanied by

a variation in the number of floral organs. There is a general significant effect towards expansion of sepal (Wilcoxon test, p-value < 2.2e-16 for MA/A and p-value < 1.3 e-16 MA/*ett-1*) and petal number (Wilcoxon test, p-value < 2.2e-4 for MA/A and p-value < 1.3 e-4 MA/*ett-1*), and reduction of stamen number (Wilcoxon test, p-value < 2.2e-16 for MA/A and p-value < 1.3 e-8 MA/*ett-1*). Moreover, sepals and petals were often narrower than in the wild-type, as has been previously described for NPA treatment of wild-type flowers in *A. thaliana*.

This experiment suggests different functional roles for two isoforms of the *ARF4* mRNA, *ARF4db* and *isoformΔ(99-170)ARF4*, in the carpel development of *Arabidopsis*. Further experiments will be necessary to directly tackle the biological function of the variant *isoformΔ(99-170)ARF4*.

Discussion

ARFs offer a good example of a family evolving by changes in coding sequence.

In a now classic paper, Marie-Claire King and Allan Wilson postulated that much of developmental evolution involves changes in gene regulation rather than molecular evolution of protein-coding sequences (26).

The evidence presented here for the evolution of novel domains in ARF proteins, and for changes affecting their overall protein structures (truncations, alternative splicings, etc.), strongly suggest that the ARF family evolves, at last partially, by changes in coding sequence.

In a general sense, we can ask whether large multigenic families do predominantly evolve by changes in coding sequence. To answer such questions, the evolution of moderate to large gene families must be investigated. In this context, the ARF family, and the Aux/IAA family whose members regulate ARF proteins, may be regarded as good candidates for study. Indeed, the Aux/IAA family is composed of 29 members in *A. thaliana*, 31 members in *O. sativa*, and 35 members in *P. trichocarpa*. Previous studies have reported some modifications in the conserved domain architecture of Aux/IAAs in flowering plants (10, 27).

The large ARF gene family as a substrate for evolution.

Within the ARF family, the rate of gene retention post-duplication seems to have been high, leading to the expansion of complexity in the ARF family. Interestingly, *ARF* genes in *P. patens* and *S. moellendorffii* seldom occur singly but rather grouped into units of several very similar genes, some of which are likely to be the result of duplication. In the case of *Selaginella*, this redundancy is not due to the sequencing of a second haplotype, since only

one haplotype was retained in the phylogenetic reconstruction. We propose that expansion of the ARF family occurred during evolution, resulting in (i) the canalization of the response to auxin during embryonic development, (ii) the diversification of ARF protein function by neofunctionalization.

The unique case of the ARF5 clade and embryo axis establishment.

MONOPTEROS (*MP*) or *ARF5* plays a role in the formation of vascular strands, and in the initiation of the body axis in early embryo in *A. thaliana*. Mutants of *mp* lack centrally located provascular cells within the basal domain of the embryo. This domain gives rise to the hypocotyl (seedling stem) and the primary root. Both of these structures are absent from mutant embryos and seedlings. Occasionally, *mp* mutants seedlings can produce adventitious roots at post-embryonic stages. The present study shows that the clade ARF5/7 is present only in seed plants, being absent in bryophytes and lycophytes. These two latter lineages have no true roots or vasculature, whose the appearance dates back to the lineage leading to pteridophytes. In lycophytes, there is evidence that these structures are secondary additions to the axis of *Lycopodium* and *Selaginella*. This is suggested by the absence of vascular tissue in the early leaves of the embryonic plant and the lack of a differentiated root-forming region in the embryo.

Moreover, the existence of vascular tissue is thought to have allowed the presence of true roots in pteridophytes. In this context, we venture the hypothesis that the clade ARF5/7 could have been determinant in the apical-basal patterning of the embryo during evolution. This hypothesis could be partially tested by cloning *ARFs* in pteridophytes.

MP shows a unique and remarkable function among ARF proteins and also contains a number of unique molecular signatures. First, ARF5 (MP) orthologues are the only ARFs lacking the conserved domain in the middle region, newly described in the paper. Second, the ARF5 clade is characterized by an absence of duplicated genes (except in *P. trichocarpa*, exhibiting two copies), whereas *O. sativa* and *P. trichocarpa* have undergone successive rounds of genome duplication (28). This observation seems analogous to the finding that genes expressed early in zebrafish development are less retained in duplicate after whole genome duplication, relative to genes expressed late (Julien Roux, personnal communication).

Regulation by *miR167* and seed plant reproduction.

Within the clade-A of ARF genes, we can distinguish a sub-clade containing ARF5 and ARF7 from one containing ARF6 and ARF8. In *A. thaliana*, *ARF6* and *ARF8* transcripts are known

to be cleavage targets of the microRNA *miR167,* which is essential for maintaining the correct expression profile of both these genes. However, there is no evidence for the regulation of *ARF5* and *ARF7* by *miR167* in *A. thaliana*. Moreover, the alignment between *miR167* and the homologous target binding site in *ARF5* shows five mismatches, whereas only two dispairs are present between *miR167* and the *ARF6* or *ARF8* transcripts. The clade A is rooted by several homologues *P. patens ARF5/7/6/8* and *S. moellendorffii ARF5/7/6/8.*

For a given miRNA, the target binding site is well conserved in the different orthologues of the miRNA-regulated gene. Moreover, most plant miRNAs have high sequence complementarity to their target binding sites, allowing a straightforward prediction of the genes they regulate (21). The target binding site of *miR167* is well conserved among seed plants, but absent in bryophytes and lycophytes. Accordingly, the *MIR167* gene is only present in seed plants and not in mosses, lycopods or ferns (29).

Towards a new consideration of alternative splicing in evolution.

In vertebrates, alternative splicing is considered to be a very important source of protein diversity, therefore emerging as a significant mechanism for the generation of regulatory diversity (30).

Most evolutionary studies of alternative splicing have used only bioinformatics data, with little knowledge of, or attention to, phenotypes. However, one study has shown a correlation between the potential of alternative splicing of glutamate receptor genes and the vocal learning capability in birds (31).

An intermediate form between ARF and Aux/IAA proteins in lower embryophytes suggests a possible evolutionary origin for the Aux/IAA family.

Aux/IAAs are known to be the main negative regulators of ARFs activity in the nucleus. In the absence of auxin, Aux/IAAs and ARFs proteins interact with each other via their domains III and IV, resulting in the repression of ARF activity. These domains are widely considered to be homologous but no evidence for this assertion has been advanced. Assuming that the C-terminal portions are homologous, two different scenarios could be taken into account : either the Aux/IAA family arose by the loss of the DNA binding domain in an ancestral ARF protein, or the ARF family arose by the addition of a DNA binding domain to domains III and IV of an ancestral Aux/IAA protein. Intuitively, several points argue in favor of the first hypothesis. First of all, Aux/IAAs regulate ARFs, therefore, logically, ARFs came first. Second, there is no evidence of ARF transcription factors lacking the DNA binding domain.

Third, we discovered a novel class of gene in *P. patens* and *S. moellendorffii*, only defined by the presence of domains III and IV. These domains are phylogenetically more closely related to domains III and IV of ARFs than to those of Aux/IAA proteins (SI Fig. 6). Moreover, canonical signatures of domains I and II of Aux/IAAs were not detected in this new class of genes. No similar forms are known in genomic databases (especially in complete genome of angiosperms), suggesting that this gene family has been lost in the lineage leading to the tracheophytes.

Materials and Methods

Plant material.

Material of *Amborella trichopoda* was field-collected from locations near Col d'Amieu, New Caledonia. Material of *Cabomba aquatica* was obtained from Anthias S.A., Les Chères 69, France. Material of *Illicium parviflorum*, *Ephedra distachya* and *Ginkgo biloba* was collected from plants cultivated in the Botanic Garden 'Parc de la Tête d'Or', Lyon, France. The *A. thaliana* ecotype Wassilewskija-2 (Ws-2) was used as the wild-type reference. Seeds of the *ett-1* mutant in the Ws-2 genetic background were obtained from the Nottingham *Arabidopsis* Stock Centre (UK). Plants were grown to maturity in peat-based compost in a growth chamber at 20°C under 16h light/8h dark cycles.

Transgenic plants.

The full *ETT*, *ARF4database* (*ARF4db*), and *isoformΔ(99-170)ARF4* coding sequences were amplified by high-fidelity PCR from a wild-type young inflorescence cDNA sample. Amplified products were inserted into the pCAMBIA3300 binary vector, which includes the *BAR* selectable marker conferring ammonium glufosinate (BASTA) resistance. Thus, the coding sequences were flanked by the *ETT* promoter (3705-bp fragment from upstream of the *Arabidopsis ETT* start codon) and the *nos* transcriptional terminator. Constructs were introduced into *Agrobacterium* strain C58pmp90 by electroporation and then into *Arabidopsis ett-1* mutant plants by floral dipping (32).

Selection of transformants.

Seeds harvested from dipped plants (T_1 seeds) were selected on MS medium containing 8 mg/ml BASTA. Resistant plants were genotyped at the endogenous *ETT* locus by using the primers 5'-CTCGATGTTAAGCTTCACG-3' and 5'-GCACTCCACCCGGTAGTGAGC-3'.

Wild-type and *ett-1/ett-1* plants were distinguished by PCR on genomic DNA, which generated 1066-bp fragment for the wild-type or the heterozygotes *ett-1/ETT* and no amplified band due to the insertion of the T-DNA for the *ett-1/ett-1* plants. We selected resistant T_1 lines that were determined to be *ett-1/ett-1* at the endogenous locus and used them for analysis in the T_2 generation during which genotypes were confirmed by monitoring Mendelian ratios.

Data collection and analysis.

ARF genes were identified in several complete genomes by tblastn using a set of selected *A. thaliana* ARFs as a probe. *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa* v1.1, *Physcomitrella patens* v1.1, and *Selaginella moellendorffii* v1.0 genomes were retrieved from JGI website (http://genome.jgi-psf.org). We also included sequences from EST databases such as NCBI, PlantGDB (33) and TIGR. For a given species, cDNA sequences were clustered and aligned using CAP3 (34) in order to identify the different alternative transcripts of a same gene.

The secondary structure of the novel motif identified in the middle region of ARFs was predicted using the MLRC method (35), available at http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sspred.html.

Molecular cloning of *ARF* cDNAs was performed in clades for which few genomic data were available. We used cDNA libraries of *A. trichopoda* and *C. aquatica* that were previously made in the laboratory (36). Other libraries were prepared using a Bacteriophage λ Uni-Zap II kit (Stratagene) from polyadenylated RNA of reproductive tissues of one species of ANA grade (*I. parviflorum*) and two gymnoperms (*E. distachya*, *G. biloba*). The screening was performed as described by Fourquin *et al.* (36).

Phylogenetic analysis.

Amino acid sequences were aligned with MUSCLE (37), manually adjusted, and selected blocks were used in the phylogenetic reconstruction.

Maximum-likelihood and bayesian inferences were performed using PhyML (38) and MrBayes (39), assuming in each case a WAG+I+Γ_4 model. Bayesian analysis was conducted for two independent runs for 5.000.000 generations and convergence was checked by plotting likelihood and parameters values upon time using Tracer 1.4 (40). The first 20% of samples were discarded as burnin and remaining trees were used to build consensus trees and calculate posterior probabilities using the 50% majority rule.

Statistical analysis.

Data were previously collected on three quantitative traits (sepal number S, petal number P, stamen number E) in *ett-1*, p*ETT*:*ARF4db*, p*ETT*:*isoformΔ(99-170)ARF4* lines of *A. thaliana*. Within each line, one hundred flowers were collected on several individuals. The variability of each trait was independently taken into account in function of the genetic background (here, our explicative variable). We estimated within-genotype (or between-individuals) variability by the non-parametric Kruskal-Wallis test given that the distributions of the number of organs are generally not normal. For each line, the effect of individual was considered as unsignificant on the variability of S, P and E (although slightly significant for E in A line) and hence pooled within line. The means of the S, P, E traits were compared by a non-parametric Wilcoxon two-sample unpaired test. All the plants were grown in the same conditions in order to discard the microenvironmental variation.

Statistical tests and graphics were performed using R statistics package (the R Project for Statistical Computing, www.r-project.org).

Acknowledgments

We thank J. Banks and A. Cuming for giving us the opportunity to respectively access the *S. moellendorfii* and *P. patens* genomes before public release. We acknowledge R. Buell, W.R. McCombie, F. Nogué and the Floral Genome Project for providing materials. We acknowledge core funding from the RDP laboratory's parent organization : CNRS, INRA, ENS-Lyon and UCBL.

References

1. Kepinski, S. & Leyser, O. (2005) *Curr Biol* **15,** R208-210.
2. Liu, C., Xu, Z., & Chua, N. H. (1993) *The Plant cell* **5,** 621-630.
3. Fischer, C. & Neuhaus, G. (1996) *Plant J* **9,** 659-669.
4. Cooke, T. J., Poli, D., Sztein, A. E., & Cohen, J. D. (2002) *Plant molecular biology* **49,** 319-338.
5. Abel, S., Oeller, P. W., & Theologis, A. (1994) *Proceedings of the National Academy of Sciences of the United States of America* **91,** 326-330.
6. Hagen, G. & Guilfoyle, T. (2002) *Plant molecular biology* **49,** 373-385.
7. Ulmasov, T., Hagen, G., & Guilfoyle, T. J. (1999) *Plant J* **19,** 309-319.
8. Tiwari, S. B., Hagen, G., & Guilfoyle, T. (2003) *The Plant cell* **15,** 533-543.
9. Ulmasov, T., Murfett, J., Hagen, G., & Guilfoyle, T. J. (1997) *The Plant cell* **9,** 1963-1971.
10. Kalluri, U. C., Difazio, S. P., Brunner, A. M., & Tuskan, G. A. (2007) *BMC Plant Biol* **7,** 59.

11. Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P. F., Lindquist, E. A., Kamisugi, Y., *et al.* (2008) *Science (New York, N.Y* **319,** 64-69.
12. Wang, D., Pei, K., Fu, Y., Sun, Z., Li, S., Liu, H., Tang, K., Han, B., & Tao, Y. (2007) *Gene* **394,** 13-24.
13. Nagpal, P., Ellis, C. M., Weber, H., Ploense, S. E., Barkawi, L. S., Guilfoyle, T. J., Hagen, G., Alonso, J. M., Cohen, J. D., Farmer, E. E., *et al.* (2005) *Development (Cambridge, England)* **132,** 4107-4118.
14. Goetz, M., Vivian-Smith, A., Johnson, S. D., & Koltunow, A. M. (2006) *The Plant cell* **18,** 1873-1886.
15. Sessions, R. A. & Zambryski, P. C. (1995) *Development (Cambridge, England)* **121,** 1519-1532.
16. Pekker, I., Alvarez, J. P., & Eshed, Y. (2005) *The Plant cell* **17,** 2899-2910.
17. Hardtke, C. S. & Berleth, T. (1998) *The EMBO journal* **17,** 1405-1411.
18. Hardtke, C. S., Ckurshumova, W., Vidaurre, D. P., Singh, S. A., Stamatiou, G., Tiwari, S. B., Hagen, G., Guilfoyle, T. J., & Berleth, T. (2004) *Development (Cambridge, England)* **131,** 1089-1100.
19. Cui, L., Wall, P. K., Leebens-Mack, J. H., Lindsay, B. G., Soltis, D. E., Doyle, J. J., Soltis, P. S., Carlson, J. E., Arumuganathan, K., Barakat, A., *et al.* (2006) *Genome Res* **16,** 738-749.
20. Williams, L., Carles, C. C., Osmont, K. S., & Fletcher, J. C. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102,** 9703-9708.
21. Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., & Bartel, D. P. (2002) *Cell* **110,** 513-520.
22. Nishimura, T., Wada, T., & Okada, K. (2004) *Biochemical Society transactions* **32,** 611-613.
23. Nishimura, T., Wada, T., Yamamoto, K. T., & Okada, K. (2005) *The Plant cell* **17,** 2940-2953.
24. Hayden, C. A. & Jorgensen, R. A. (2007) *BMC biology* **5,** 32.
25. Guilfoyle, T. & Hagen, G. (2001) *Journal of Plant Growth Regulation* **10,** 281-291.
26. King, M. C. & Wilson, A. C. (1975) *Science (New York, N.Y* **188,** 107-116.
27. Jain, M., Kaur, N., Garg, R., Thakur, J. K., Tyagi, A. K., & Khurana, J. P. (2006) *Functional & integrative genomics* **6,** 47-59.
28. Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., *et al.* (2006) *Science (New York, N.Y* **313,** 1596-1604.
29. Axtell, M. J. & Bartel, D. P. (2005) *The Plant cell* **17,** 1658-1673.
30. Black, D. L. (2000) *Cell* **103,** 367-370.
31. Wada, K., Sakaguchi, H., Jarvis, E. D., & Hagiwara, M. (2004) *The Journal of comparative neurology* **476,** 44-64.
32. Clough, S. J. & Bent, A. F. (1998) *Plant J* **16,** 735-743.
33. Dong, Q., Schlueter, S. D., & Brendel, V. (2004) *Nucleic acids research* **32,** D354-359.
34. Huang, X. & Madan, A. (1999) *Genome Research* **9,** 868-877.
35. Guermeur, Y., Geourjon, C., Gallinari, P., & Deleage, G. (1999) *Bioinformatics (Oxford, England)* **15,** 413-421.
36. Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C., & Scutt, C. P. (2005) *Proceedings of the National Academy of Sciences of the United States of America* **102,** 4649-4654.
37. Edgar, R. C. (2004) *BMC bioinformatics* **5,** 113.

38. Guindon, S. & Gascuel, O. (2003) *Systematic biology* **52,** 696-704.
39. Ronquist, F. & Huelsenbeck, J. P. (2003) *Bioinformatics (Oxford, England)* **19,** 1572-1574.
40. Rambaut, A. & Drummond, A. J. (2007).

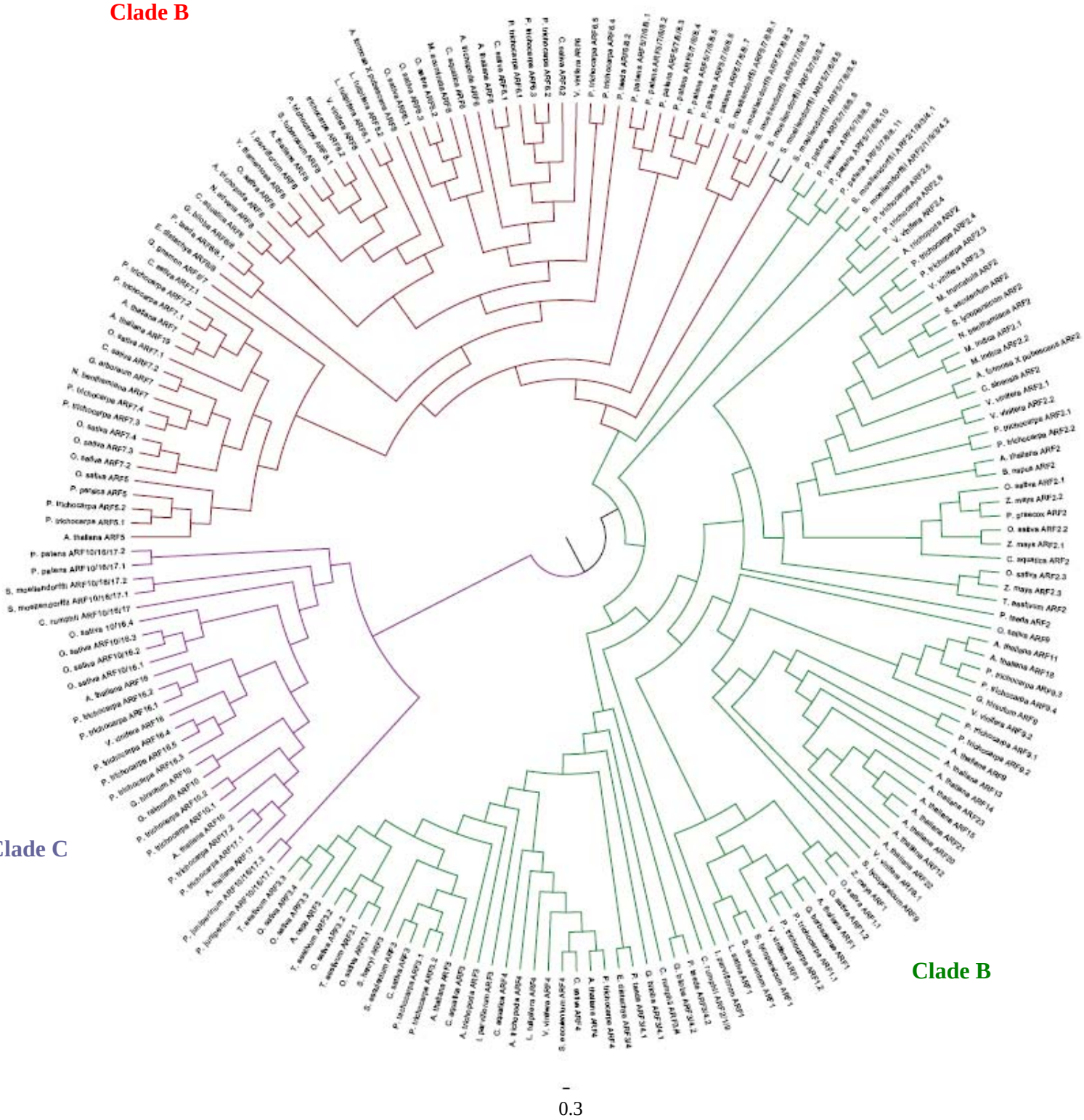


Figure 1. Phylogenetic tree of *Auxin Response Factor* genes in the land plants. Bayesian phylogram is based on the amino-acid sequences. ML analysis yielded similar results (see text). The clade A is indicated in red, the clade B in green and the clade C in violet.

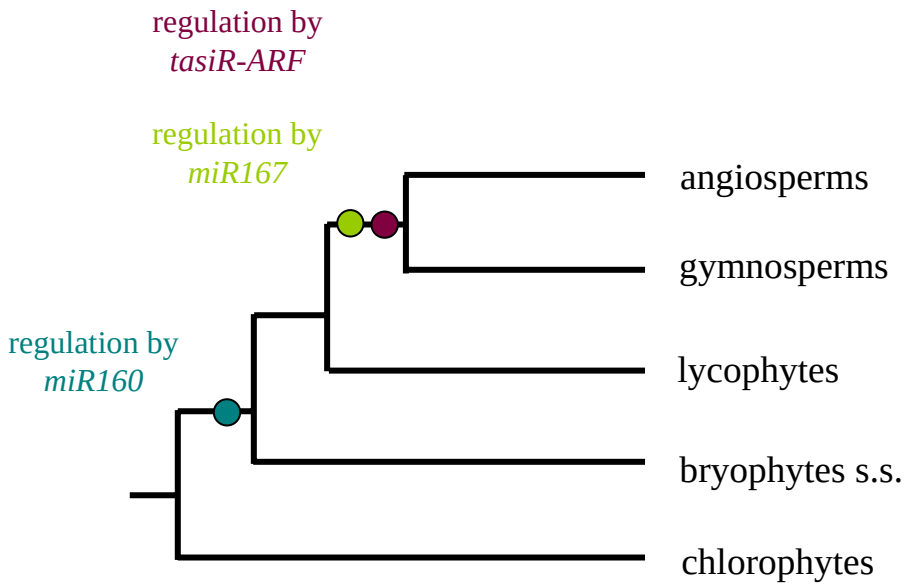


Figure 2. Phylogenetic distribution of the reported regulations of ARFs.

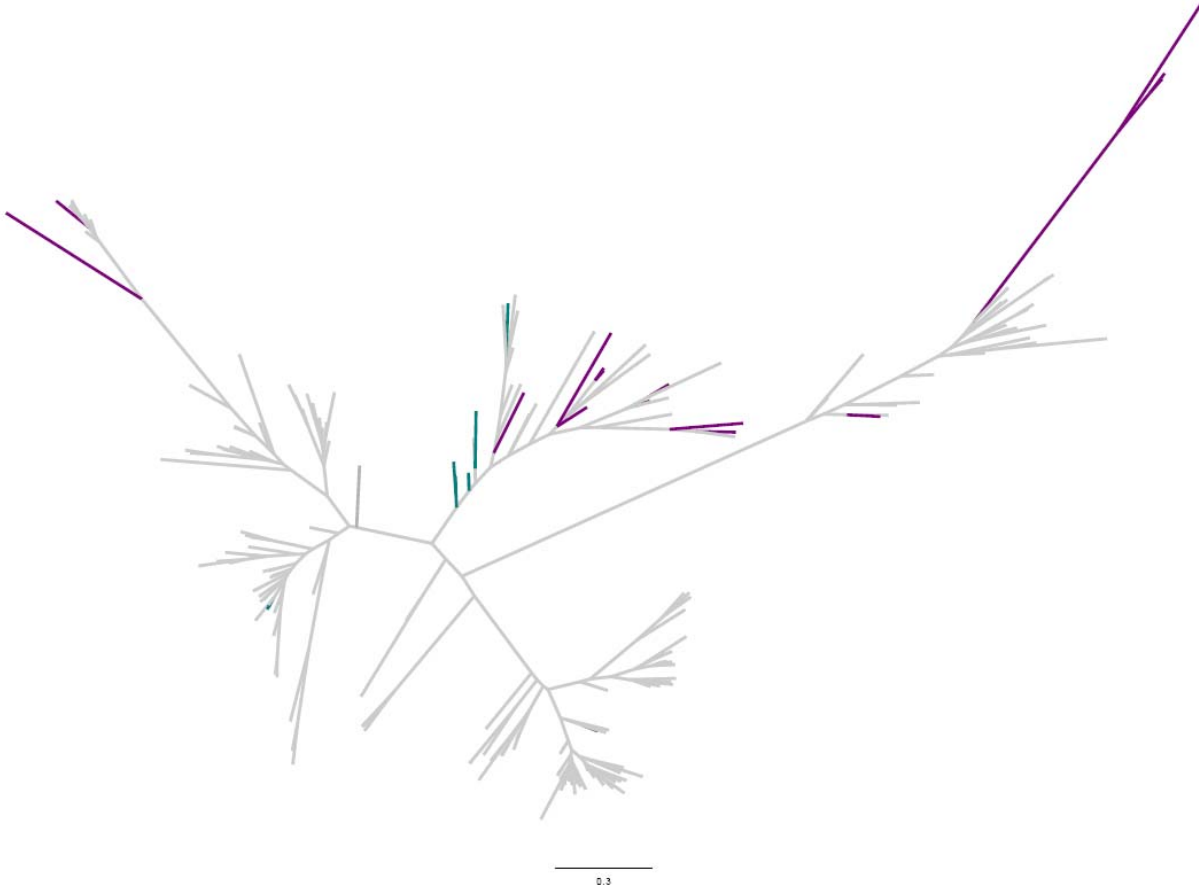
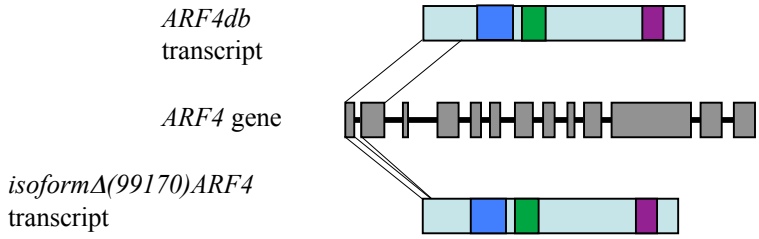
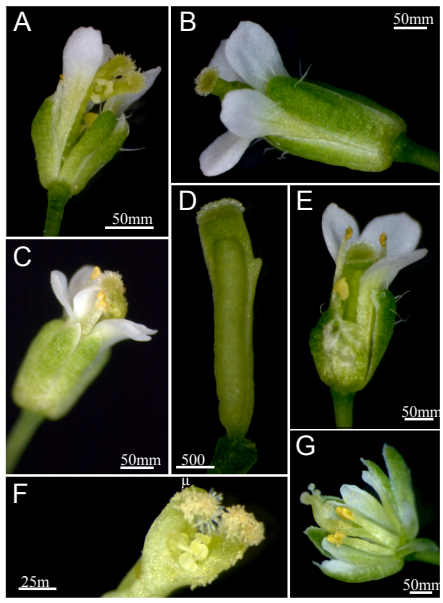


Figure 3. Phylogenetic distribution of truncations and alternative splicings. Events of truncation at the genomic level are colored violet. Events of alternative splicing are colored green.





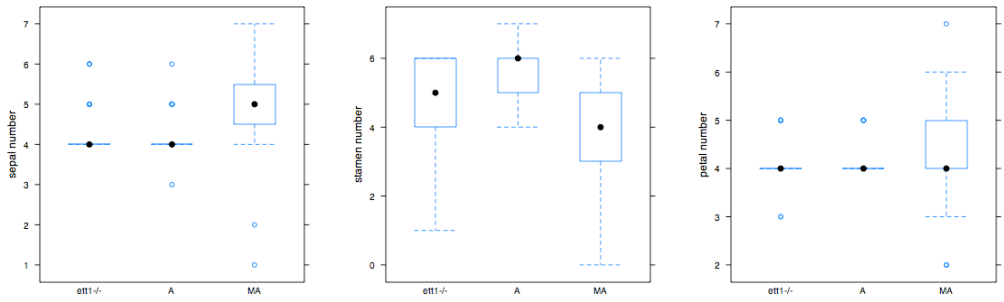


Figure 4. Complementation of the *ett-1* mutation by transformation with two isoforms of *AtARF4* under the control of *ARF3* promoter.

a. Schematic representation of the two isoforms. b. Floral phenotypes of *ett-1* mutants (A), *ett-1* mutants transformed with the coding sequence of *ARF3* (B), *ARF4db* (D, E) and *isoformΔ(99-170)ARF4* (C,F,G). c. Boxplots illustrating the variation in organ numbers in the different lines.

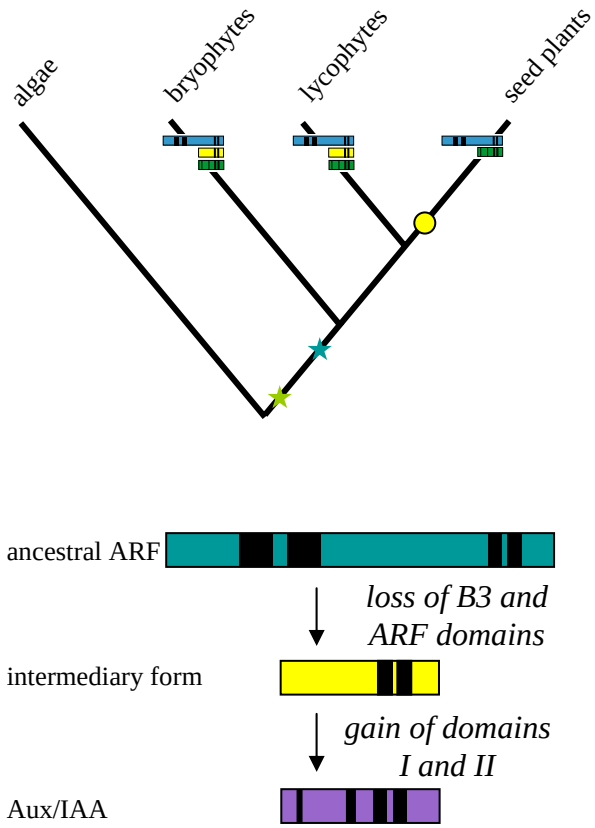
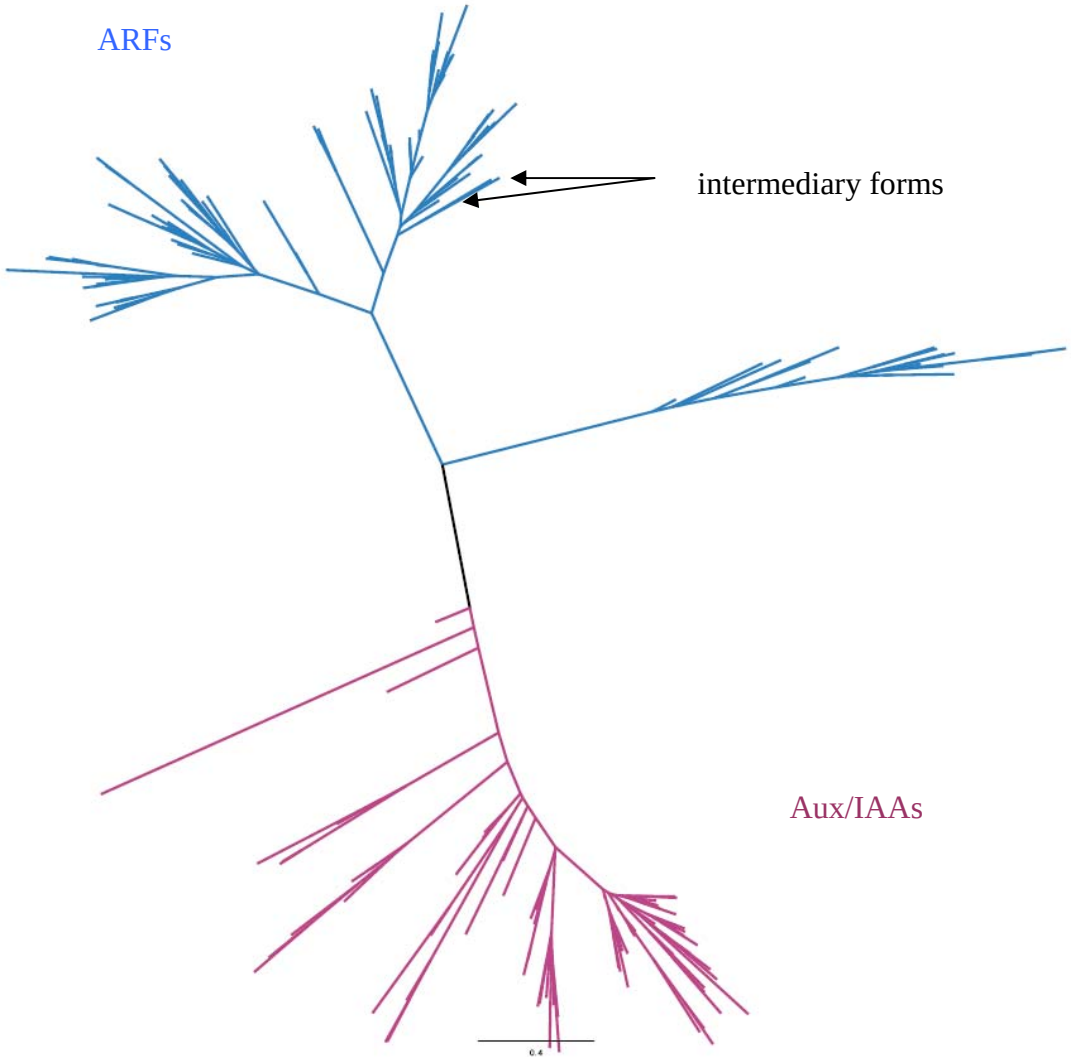


Figure 5. An evolutionary hypothesis for the appearance of the regulation of ARFs by Aux/IAAs. Our results suggest that the transcriptional regulators Aux/IAA could have been derived from ARF proteins during the evolution.



SI Figure 6. Phylogenetic tree of ARFs and Aux/IAAs based on domains III and IV. ARF proteins are in blue and Aux/IAA proteins are in pink. Arrows indicate the phylogenetic position of the two proteins predicted in the genome of *P. patens* and *S. moellendorffii* which are characterized by the absence of B3 and ARF domains.

Species	Name	Accession number
Allium cepa	*A. cepa ARF3*	TC1453
Amborella trichopoda	*A. trichopoda ARF2*	
Amborella trichopoda	*A. trichopoda ARF3*	
Amborella trichopoda	*A. trichopoda ARF4*	
Amborella trichopoda	*A. trichopoda ARF6*	
Amborella trichopoda	*A. trichopoda ARF8*	
Aquilegia formosa X pubescens	*A. formosa X pubescens ARF2*	PUT-157a-12123
Aquilegia formosa X pubescens	*A. formosa X pubescens ARF8*	TC8454
Arabidopsis thaliana	*A. thaliana ARF1*	NP_176184
Arabidopsis thaliana	*A. thaliana ARF2*	NP_851244
Arabidopsis thaliana	*A. thaliana ARF3*	NP_180942
Arabidopsis thaliana	*A. thaliana ARF4*	NP_200853
Arabidopsis thaliana	*A. thaliana ARF5*	NP_173414
Arabidopsis thaliana	*A. thaliana ARF6*	NP_174323
Arabidopsis thaliana	*A. thaliana ARF7*	NP_851047
Arabidopsis thaliana	*A. thaliana ARF8*	NP_198518
Arabidopsis thaliana	*A. thaliana ARF9*	NP_194129
Arabidopsis thaliana	*A. thaliana ARF10*	NP_180402
Arabidopsis thaliana	*A. thaliana ARF11*	NP_182176
Arabidopsis thaliana	*A. thaliana ARF12*	NP_174691
Arabidopsis thaliana	*A. thaliana ARF13*	
Arabidopsis thaliana	*A. thaliana ARF14*	NP_174786
Arabidopsis thaliana	*A. thaliana ARF15*	Q9LQE3
Arabidopsis thaliana	*A. thaliana ARF16*	NP_567841
Arabidopsis thaliana	*A. thaliana ARF17*	NP_565161
Arabidopsis thaliana	*A. thaliana ARF18*	NP_567119
Arabidopsis thaliana	*A. thaliana ARF19*	NP_173356
Arabidopsis thaliana	*A. thaliana ARF20*	
Arabidopsis thaliana	*A. thaliana ARF21*	NP_174701
Arabidopsis thaliana	*A. thaliana ARF22*	NP_174699
Arabidopsis thaliana	*A. thaliana ARF23*	
Brassica napus	*B. napus ARF2*	AJ716227
Cabomba aquatica	*C. aquatica ARF2*	
Cabomba aquatica	*C. aquatica ARF3*	
Cabomba aquatica	*C. aquatica ARF4*	
Cabomba aquatica	*C. aquatica ARF6*	
Cabomba aquatica	*C. aquatica ARF8*	
Citrus sinensis	*C. sinensis ARF2*	PUT-157a-13256
Cucumis sativa	*C. sativa ARF3*	CO995305
Cucumis sativa	*C. sativa ARF4*	BAD19065
Cucumis sativa	*C. sativa ARF6.1*	BAD19063
Cucumis sativa	*C. sativa ARF6.2*	BAD19064
Cucumis sativa	*C. sativa ARF7.1*	BAD19061
Cucumis sativa	*C. sativa ARF7.2*	BAD19062
Cycas rumphii	*C. rumphii ARF2/1/9*	
Cycas rumphii	*C. rumphii ARF3/4*	
Cycas rumphii	*C. rumphii ARF10/16/17*	
Ephedra distachya	*E. distachya ARF3/4*	
Ephedra distachya	*E. distachya ARF6/8*	
Ginkgo biloba	*G. biloba ARF3/4.1*	
Ginkgo biloba	*G. biloba ARF3/4.2*	
Ginkgo biloba	*G. biloba ARF6/8*	
Gnetum gnemon	*G. gnemon ARF5/7*	

Gossypium arboreum	*G. arboreum ARF7*	AAZ81522
Gossypium barbadense	*G. barbadense ARF1*	AAZ81521
Gossypium hirsutum	*G. hirsutum ARF9*	TC21633
Gossypium hirsutum	*G. hirsutum ARF10*	ABO60876
Gossypium raimondii	*G. raimondii ARF10*	AAX89755
Illicium parviflorum	*I. parviflorum ARF1*	
Illicium parviflorum	*I. parviflorum ARF3*	
Illicium parviflorum	*I. parviflorum ARF8*	
Lactuca sativa	*L. sativa ARF1*	BU013963
Liriodendron tulipifera	*L. tulipifera ARF4*	DT584664
Liriodendron tulipifera	*L. tulipifera ARF8.1*	DT587096
Liriodendron tulipifera	*L. tulipifera ARF8.2*	DT595650
Mangifera indica	*M. indica ARF2.1*	AAP06759
Mangifera indica	*M. indica ARF2.2*	AAP57471
Medicago truncatula	*M. truncatula ARF2*	TC106929
Musa acuminata	*M. acuminata ARF6*	ABF69979
Nicotiana benthamiana	*N. benthamiana ARF2*	TC9336
Nicotiana benthamiana	*N. benthamiana ARF7*	TC9737
Nuphar advena	*N. advena ARF8*	CD476221
Oryza sativa	*O. sativa ARF1.1*	NM_001059414
Oryza sativa	*O. sativa ARF1.2*	NM_001053672
Oryza sativa	*O. sativa ARF2.1*	NM_001073297
Oryza sativa	*O. sativa ARF2.2*	NM_001074520
Oryza sativa	*O. sativa ARF2.3*	NM_001051805
Oryza sativa	*O. sativa ARF3.1*	NM_001050360
Oryza sativa	*O. sativa ARF3.2*	AB071290
Oryza sativa	*O. sativa ARF3.3*	AK067927
Oryza sativa	*O. sativa ARF3.4*	NM_001050809
Oryza sativa	*O. sativa ARF5*	
Oryza sativa	*O. sativa ARF6.1*	NM_001073297
Oryza sativa	*O. sativa ARF6.2*	NM_001064895
Oryza sativa	*O. sativa ARF6.3*	NM_001052528
Oryza sativa	*O. sativa ARF7.1*	NM_001063589
Oryza sativa	*O. sativa ARF7.2*	NM_001068803
Oryza sativa	*O. sativa ARF7.3*	
Oryza sativa	*O. sativa ARF7.4*	
Oryza sativa	*O. sativa ARF8*	NM_001060757
Oryza sativa	*O. sativa ARF9*	NM_001049063
Oryza sativa	*O. sativa ARF10/16.1*	NM_001071407
Oryza sativa	*O. sativa ARF10/16.2*	AK243230
Oryza sativa	*O. sativa ARF10/16.3*	Q01I35
Oryza sativa	*O. sativa ARF10/16.4*	NM_001064934
Phyllostachys praecox	*P. praecox ARF2*	DQ013802
Physcomitrella patens	*P. patens ARF5/7/6/8.1*	
Physcomitrella patens	*P. patens ARF5/7/6/8.2*	
Physcomitrella patens	*P. patens ARF5/7/6/8.3*	estExt_Genewise1.C_650154
Physcomitrella patens	*P. patens ARF5/7/6/8.4*	
Physcomitrella patens	*P. patens ARF5/7/6/8.5*	gw1.119.8.1
Physcomitrella patens	*P. patens ARF5/7/6/8.6*	
Physcomitrella patens	*P. patens ARF5/7/6/8.7*	
Physcomitrella patens	*P. patens ARF5/7/6/8.8*	
Physcomitrella patens	*P. patens ARF5/7/6/8.9*	
Physcomitrella patens	*P. patens ARF5/7/6/8.10*	gw1.280.8.1
Physcomitrella patens	*P. patens ARF5/7/6/8.11*	GW1.341.65.1
Physcomitrella patens	*P. patens ARF10/16/17.1*	FGENESH1_PM.SCAFFOLD_339000005

Physcomitrella patens	*P. patens ARF10/16/17.2*	E_GW1.279.55.1
Pinus taeda	*P. taeda ARF2*	TC70680
Pinus taeda	*P. taeda ARF3/4.1*	TC64514
Pinus taeda	*P. taeda ARF3/4.2*	BX250119
Pinus taeda	*P. taeda ARF6/8.1*	TC75609
Pinus taeda	*P. taeda ARF6/8.2*	TC65336
Polytrichum juniperinum	*P. juniperinum ARF10/16/17.1*	AAY17065
Polytrichum juniperinum	*P. juniperinum ARF10/16/17.2*	AAY17064
Populus trichocarpa	*P. trichocarpa ARF1.1*	estExt_fgenesh4_pg.C_1500013
Populus trichocarpa	*P. trichocarpa ARF1.2*	estExt_fgenesh4_pm.C_860029
Populus trichocarpa	*P. trichocarpa ARF2.1*	estExt_fgenesh4_pm.C_LG_XII0386
Populus trichocarpa	*P. trichocarpa ARF2.2*	eugene3.00150845
Populus trichocarpa	*P. trichocarpa ARF2.3*	gw1.I.7907.1
Populus trichocarpa	*P. trichocarpa ARF2.4*	estExt_fgenesh4_pm.C_LG_III0550
Populus trichocarpa	*P. trichocarpa ARF2.5*	gw1.XIV.424.1
Populus trichocarpa	*P. trichocarpa ARF2.6*	gw1.XIV.2727.1
Populus trichocarpa	*P. trichocarpa ARF3.1*	estExt_Genewise1_v1.C_LG_IV2935
Populus trichocarpa	*P. trichocarpa ARF3.2*	fgenesh4_pg.C_scaffold_187000006
Populus trichocarpa	*P. trichocarpa ARF4*	fgenesh4_pg.C_LG_IX001438
Populus trichocarpa	*P. trichocarpa ARF5.1*	grail3.0003020302
Populus trichocarpa	*P. trichocarpa ARF5.2*	grail3.0002064402
Populus trichocarpa	*P. trichocarpa ARF6.1*	fgenesh4_pg.C_LG_I002802
Populus trichocarpa	*P. trichocarpa ARF6.2*	estExt_Genewise1_v1.C_LG_XI2869
Populus trichocarpa	*P. trichocarpa ARF6.3*	fgenesh4_pg.C_scaffold_1006000001
Populus trichocarpa	*P. trichocarpa ARF6.4*	gw1.II.3332.1
Populus trichocarpa	*P. trichocarpa ARF6.5*	gw1.V.808.1
Populus trichocarpa	*P. trichocarpa ARF7.1*	fgenesh4_pg.C_LG_XVIII000045
Populus trichocarpa	*P. trichocarpa ARF7.2*	eugene3.00280060
Populus trichocarpa	*P. trichocarpa ARF7.3*	estExt_fgenesh4_pg.C_1640064
Populus trichocarpa	*P. trichocarpa ARF7.4*	gw1.VI.2366.1
Populus trichocarpa	*P. trichocarpa ARF8.1*	gw1.IV.3880.1
Populus trichocarpa	*P. trichocarpa ARF8.2*	gw1.44.432.1
Populus trichocarpa	*P. trichocarpa ARF9.1*	gw1.III.1140.1
Populus trichocarpa	*P. trichocarpa ARF9.2*	gw1.I.8521.1
Populus trichocarpa	*P. trichocarpa ARF9.3*	gw1.II.1258.1
Populus trichocarpa	*P. trichocarpa ARF9.4*	gw1.XIV.1750.1
Populus trichocarpa	*P. trichocarpa ARF10.1*	eugene3.00660262
Populus trichocarpa	*P. trichocarpa ARF10.2*	fgenesh4_pg.C_LG_IX001411
Populus trichocarpa	*P. trichocarpa ARF16.1*	fgenesh4_pg.C_LG_VIII000301
Populus trichocarpa	*P. trichocarpa ARF16.2*	eugene3.00080331
Populus trichocarpa	*P. trichocarpa ARF16.3*	estExt_fgenesh4_pm.C_LG_XVI0323
Populus trichocarpa	*P. trichocarpa ARF16.4*	gw1.28.631.1
Populus trichocarpa	*P. trichocarpa ARF16.5*	gw1.28.632.1
Populus trichocarpa	*P. trichocarpa ARF17.1*	eugene3.00020832
Populus trichocarpa	*P. trichocarpa ARF17.2*	estExt_fgenesh4_pg.C_LG_V0901
Prunus persica	*P. persica ARF5*	AAO14628
Saruma henryi	*S. henryi ARF3*	DT590233
Selaginella moellendorffii	*S. moellendorffii ARF5/7/6/8.1*	estExt_Genewise1Plus.C_650169
Selaginella moellendorffii	*S. moellendorffii ARF5/7/6/8.2*	E_GW1.72.101.1
Selaginella moellendorffii	*S. moellendorffii ARF5/7/6/8.3*	FGENESH2_PG.C_SCAFFOLD_65000062
Selaginella moellendorffii	*S. moellendorffii ARF5/7/6/8.4*	GW1.72.74.1
Selaginella moellendorffii	*S. moellendorffii ARF5/7/6/8.5*	GW1.113.54.1
Selaginella moellendorffii	*S. moellendorffii ARF5/7/6/8.6*	E_GW1.113.82.1
Selaginella moellendorffii	*S. moellendorffii ARF2/1/9/3/4.1*	E_GW1.127.60.1
Selaginella moellendorffii	*S. moellendorffii ARF2/1/9/3/4.2*	ESTEXT_FGENESH2_PG.C_10526

Selaginella moellendorffii	*S. moellendorffii ARF10/16/17.1*	GW1.45.343.1
Selaginella moellendorffii	*S. moellendorffii ARF10/16/17.2*	GW1.0.1610.1
Solanum esculentum	*S. esculentum ARF1*	TC185437
Solanum esculentum	*S. esculentum ARF2*	TC170930
Solanum esculentum	*S. esculentum ARF3*	TC175247
Solanum esculentum	*S. esculentum ARF4*	TC188486
Solanum lycopersicum	*S. lycopersicum ARF1*	TC32883
Solanum lycopersicum	*S. lycopersicum ARF2*	TC48368
Solanum lycopersicum	*S. lycopersicum ARF9*	BT013639
Solanum tuberosum	*S. tuberosum ARF8*	TC156978 + CX700142
Triticum aestivum	*T. aestivum ARF2*	AAW82475
Triticum aestivum	*T. aestivum ARF3.1*	AAQ86958
Triticum aestivum	*T. aestivum ARF3.2*	AAQ86959
Triticum aestivum	*T. aestivum ARF3.3*	AAQ86960
Vitis vinifera	*V. vinifera ARF1*	CAO46911
Vitis vinifera	*V. vinifera ARF2.1*	CAO68379
Vitis vinifera	*V. vinifera ARF2.2*	CAN67897
Vitis vinifera	*V. vinifera ARF2.3*	TC70940
Vitis vinifera	*V. vinifera ARF2.4*	CAO22423
Vitis vinifera	*V. vinifera ARF4*	CAO21817
Vitis vinifera	*V. vinifera ARF6*	CAO40364
Vitis vinifera	*V. vinifera ARF8*	CAO22179
Vitis vinifera	*V. vinifera ARF9.1*	CAO69513
Vitis vinifera	*V. vinifera ARF9.2*	CAO14392
Vitis vinifera	*V. vinifera ARF16*	TC71158
Yucca filamentosa	*Y. filamentosa ARF8*	DT581114
Zea mays	*Z. mays ARF1*	TC318047
Zea mays	*Z. mays ARF2.1*	TC66226
Zea mays	*Z. mays ARF2.2*	TC333094
Zea mays	*Z. mays ARF2.3*	TC325773

SI Table 1. Database accession numbers of genes used for the phylogenetic reconstruction.

	5' UTR	uORFs	ta-siRNA	miR160	QSL-rich	new site	miR167
A. thaliana ARF5	474 bp	8					
P. trichocarpa ARF5.1	9 bp						
P. trichocarpa ARF5.2	9 bp						
P. persica ARF5							
O. sativa ARF5	532 bp	1					
O. sativa ARF7.2	135 bp	1					
O. sativa ARF7.3	436 bp	3					
O. sativa ARF7.4							
P. trichocarpa ARF7.3							
P. trichocarpa ARF7.4							
N. benthamiana ARF7	472 bp	2					
G. arboreum ARF7	112 bp	2					
C. sativa ARF7.2							
O. sativa ARF7.1							
A. thaliana ARF19	308 bp	2					
A. thaliana ARF7	609 bp	2					
P. trichocarpa ARF7.1							
P. trichocarpa ARF7.2							
C. sativa ARF7.1	700 bp	3					
G. gnemon ARF5/7							
E. distachya ARF6/8	58 bp	1					
P. taeda ARF6/8.1							
G. biloba ARF6/8	424 bp	6					
C. aquatica ARF8							
N. advena ARF8							
A. trichopoda ARF8	71 bp	2					
O. sativa ARF8	353 bp	4					
Y. filamentosa ARF8	231 bp	3					
I. parviflorum ARF8	343 bp	8					
A. thaliana ARF8	752 bp	10					
S. tuberosum ARF8							
P. trichocarpa ARF8.1							
P. trichocarpa ARF8.2							
V. vinifera ARF8							
L. tulipifera ARF8.1							
L. tulipifera ARF8.2							
A. formosa X pubescens ARF8							
O. sativa ARF6.1	881 bp	8					
O. sativa ARF6.3	758 bp	6					
O. sativa ARF6.2	696 bp	3					
M. acuminata ARF6							
C. aquatica ARF6							
A. trichopoda ARF6	191 bp	1					
A. thaliana ARF6	763 bp	7					
C. sativa ARF6.1	504 bp	5					
P. trichocarpa ARF6.1							
P. trichocarpa ARF6.3							
P. trichocarpa ARF6.2							
C. sativa ARF6.2							
V. vinifera ARF6							
P. trichocarpa ARF6.5							
P. trichocarpa ARF6.4							
P. taeda ARF6/8.2							
P. patens ARF5/7/6/8.1							
P. patens ARF5/7/6/8.2							

	5' UTR	uORFs	ta-siRNA	miR160	QSL-rich	new site	miR167
P. patens ARF5/7/6/8.3							
P. patens ARF5/7/6/8.4							
P. patens ARF5/7/6/8.5							
P. patens ARF5/7/6/8.6							
P. patens ARF5/7/6/8.7							
S. moellendorffii ARF5/7/6/8.1							
S. moellendorffii ARF5/7/6/8.2							
S. moellendorffii ARF5/7/6/8.3							
S. moellendorffii ARF5/7/6/8.4							
S. moellendorffii ARF5/7/6/8.5							
S. moellendorffii ARF5/7/6/8.6							
P. patens ARF2/1/9/3/4.1							
P. patens ARF2/1/9/3/4.2							
P. patens ARF2/1/9/3/4.3							
P. patens ARF2/1/9/3/4.4							
S. moellendorffii ARF2/1/9/3/4.1							
S. moellendorffii ARF2/1/9/3/4.2							
P. trichocarpa ARF2.5							
P. trichocarpa ARF2.6							
V. vinifera ARF2.4							
A. trichopoda ARF2							
P. trichopoda ARF2.4							
P. trichopoda ARF2.3							
V. vinifera ARF2.3							
M. truncatula ARF2	348 bp	1					
S. esculentum ARF2	125 bp	1					
S. lycopersicum ARF2	125 bp	1					
N. benthamiana ARF2	108 bp	1					
M. indica ARF2.1	285 bp	3					
M. indica ARF2.2	285 bp	3					
A. formosa X pubescens ARF2							
C. sinensis ARF2	123 bp	1					
V. vinifera ARF2.1							
V. vinifera ARF2.2							
P. trichocarpa ARF2.1							
P. trichocarpa ARF2.2							
A. thaliana ARF2	501 bp	2					
B. napus ARF2							
O. sativa ARF2.1	332 bp	1					
Z. mays ARF2.2	269 bp	2					
P. praecox ARF2	21 bp	1					
O. sativa ARF2.2	210 bp	1					
Z. mays ARF2.1							
C. aquatica ARF2							
O. sativa ARF2.3	224 bp						
Z. mays ARF2.3	177 bp						
T. aestivum ARF2							
P. taeda ARF2							
O. sativa ARF9	194 bp						
A. thaliana ARF11	328 bp	7					
A. thaliana ARF18	282 bp	2					
P. trichopoda ARF9.3							
P. trichopoda ARF9.4							
G. hirsutum ARF9							

	5' UTR	uORFs	ta-siRNA	miR160	QSL-rich	new site	miR167
V. vinifera ARF9.2							
P. trichopoda ARF9.1							
P. trichopoda ARF9.2							
A. thaliana ARF9	315 bp	1					
A. thaliana ARF13							
A. thaliana ARF14							
A. thaliana ARF23							
A. thaliana ARF15							
A. thaliana ARF21							
A. thaliana ARF20							
A. thaliana ARF12							
A. thaliana ARF22							
V. vinifera ARF9.1							
S. lycopersicum ARF9	120 bp	1					
Z. mays ARF1							
O. sativa ARF1.1	573 bp	3					
O. sativa ARF1.2	300 bp	1					
A. thaliana ARF1	266 bp						
G. barbadense ARF1							
P. trichocarpa ARF1.1	99 bp						
P. trichocarpa ARF1.2							
V. vinifera ARF1							
S. lycopersicon ARF1	348 bp	3					
S. esculentum ARF1	348 bp	3					
L. sativa ARF1	216 bp	1					
I. parviflorum ARF1	678 bp	4					
C. rumphii ARF2/1/9							
P. taeda ARF3/4.2	250 bp	2					
G. biloba ARF3/4.2							
C. rumphii ARF3/4							
G. biloba ARF3/4.1							
P. taeda ARF3/4.1							
E. distachya ARF3/4	390 bp	6					
P. trichocarpa ARF4							
A. thaliana ARF4	511 bp	4					
C. sativa ARF4	122 bp						
S. esculentum ARF4	464 bp	6					
V. vinifera ARF4							
L. tulipifera ARF4							
A. trichopoda ARF4	115 bp	1					
C. aquatica ARF4	66 bp						
I. parviflorum ARF3	187 bp	1					
A. trichopoda ARF3	35 bp						
C. aquatica ARF3	194 bp	7					
A. thaliana ARF3	365 bp	2					
P. trichopoda ARF3.2							
P. trichopoda ARF3.1							
C. sativa ARF3							
S. esculentum ARF3	45 bp						
S. henryi ARF3							
O. sativa ARF3.1	157 bp						
T. aestivum ARF3.1							
O. sativa ARF3.2	221 bp						
T. aestivum ARF3.2							
A. cepa ARF3	118 bp						

	5' UTR	uORFs	ta-siRNA	miR160	QSL-rich	new site	miR167
O. sativa ARF3.3	315 bp	1					
O. sativa ARF3.4	293 bp	3					
T. aestivum ARF3.3							
P. juniperinum ARF10/16/17.1							
P. juniperinum ARF10/16/17.2							
A. thaliana ARF17							
P. trichocarpa ARF17.1							
P. trichocarpa ARF17.2	38 bp						
A. thaliana ARF10							
P. trichocarpa ARF10.1							
P. trichocarpa ARF10.2							
G. raimondii ARF10	77 bp						
G. hirsutum ARF10	77 bp						
P. trichocarpa ARF16.3							
P. trichocarpa ARF16.5							
P. trichocarpa ARF16.4							
V. vinifera ARF16							
P. trichocarpa ARF16.1							
P. trichocarpa ARF16.2							
A. thaliana ARF16							
O. sativa ARF10/16.1	276 bp	1					
O. sativa ARF10/16.2	155 bp						
O sativa ARF10/16.3							
O. sativa 10/16.4	666 bp	6					
C. rumphii ARF10/16/17							
S. moellendorffii ARF10/16/17.1							
S. moellendorffii ARF10/16/17.2							
P. patens ARF10/16/17.1							
P. patens ARF10/16/17.2							

SI Table 2. Presence/absence of the signatures of conserved regulatory sites for each sequence used in this study. Black, grey, and white boxes respectively indicate presence, lack of data, and absence of a regulatory domain.

III) Perspectives

L'échantillonnage effectué dans l'approche ci-dessus est assez bien représentatif des embryophytes. Cependant, des embranchements entiers restent sous représentés comme le sont les hépaticophytes, les anthocérotophytes et les filicophytes. Concernant les bryophytes *senso latu*, la seule manière de combler le biais en matière de données génomiques serait d'envisager de construire des banques d'ESTs. Pour ce qui est des filicophytes, il ne serait pas étonnant que l'espèce *Ceratopteris richardii* fasse le sujet d'un projet de séquençage dans les prochaines années. En effet, de nombreux ESTs sont déjà disponibles chez cette espèce et chose rare à souligner : la transformation génétique semble être possible chez cette espèce (Rutherford et al., 2004).

Chez *A. thaliana*, l'étude des gènes *ARFs* et *Aux/IAAs* se révèle ardue dans le sens où une forte redondance génétique existe entre les différents membres d'une même famille. Une manière possible de simplifier le modèle est d'étudier quels sont les gènes (*ARFs* et *Aux/IAAs*) qui s'expriment au même endroit (définition tissulaire, voire cellulaire) à un instant t du développement. Ainsi, on passe d'un grand nombre d'interactions possibles testées biochimiquement à un nombre plus restreint de combinaisons ayant réellement lieu *in planta* (projet mené dans le laboratoire par Dr. Teva Vernoux). Une autre manière d'envisager les choses serait de se placer chez *P. patens* ou *S. moellendorffii* chez lesquelles les gènes *ARFs* et *Aux/IAAs* sont moins dupliqués. Puis, ayant retracé l'histoire évolutive de ces gènes chez les embryophytes, il sera possible d'inférer des hypothèses quant au mode de fonctionnement de la boucle de régulation ARF-Aux/IAA chez l'espèce modèle *A. thaliana*.

Cette étude a permis de mettre en évidence la présence d'un motif hautement conservé dans la région centrale des protéines ARFs. À l'exception du clade ARF5, ce motif est présent chez l'ensemble des membres de la famille. Des recherches *in silico* de motifs protéiques similaires déjà connus n'ont donné aucun résultat. Il faudrait envisager une approche expérimentale pour appréhender la fonction moléculaire de ce motif. D'un point de vue *in vitro*, il serait intéressant d'introduire des mutations ponctuelles dans ce motif (mutagenèse dirigée) et de caractériser l'effet de ces mutations sur les fonctions des ARFs. Par exemple, on peut envisager que ce motif intervienne d'une manière ou d'une autre sur la capacité de dimérisation des ARFs. Une manière de tester cette hypothèse serait de réaliser des manipulations de type EMSA (**E**lectrophoretic **M**obility **S**hift **A**ssay) entre des protéines ARFs et des séquences nucléotidiques AuxREs, sachant que la stabilité de l'interaction diffère selon le degré de dimérisation des ARFs (Ulmasov et al., 1999). Un deuxième rôle possible de ce motif serait qu'il intervient dans l'activité activatrice ou inhibitrice du domaine central.

Cette hypothèse pourrait être testée en (i) fusionnant la région centrale des ARFs comprenant le motif muté avec le domaine de liaison à l'ADN de GAL4 (GAL4-DBD) de la levure, (ii) testant ces construction chimériques par transfection transitoire de protoplastes avec un promoteur constitutif ou minimal GUS contenant un site de liaison à GAL4, (iii) comparant l'activation dans un contexte sauvage ou muté pour le motif nouvellement décrit.

Chapitre 4 : GÈNES *YABBY* ET ÉVOLUTION DE LA FLEUR

I) Introduction et résumé de l'article

Les gènes *YABBY* jouent un rôle central dans l'établissement de la polarité adaxiale/abaxiale des organes latéraux chez les angiospermes. En particulier, les gènes *CRABS CLAW* (*CRC*) et *INNER NO OUTER* (*INO*) interviennent respectivement dans la mise en place du caractère abaxial du carpelle et de l'ovule chez *A. thaliana*. De plus, leurs expressions restreintes au gynécée et aux nectaires pour *CRC* et aux ovules pour *INO* en font d'excellents marqueurs de la fleur.

Le but de cet article est de reconstruire l'histoire évolutive de cette famille multigénique (6 membres chez *A. thaliana*) et de tester l'hypothèse "Mostly Male" à partir des données d'expression des orthologues du gène *INO* chez les gymnospermes. Cette analyse se restreint à l'étude des gymnospermes dans le sens où la phylogénie des gènes *YABBY* chez les angiospermes est maintenant bien documentée (Lee et al., 2005 ; Toriba et al., 2007). Elle repose sur plusieurs étapes : (i) le clonage exhaustif de gènes *YABBY* chez un grand nombre de gymnospermes, échantillonnés au Jardin Botanique du Parc de la Tête d'Or (Lyon), (ii) la reconstruction phylogénétique de cette famille chez les gymnospermes, (iii) la caractérisation de leur patron d'expression.

J'ai contribué à ce projet en différents points. J'ai tout d'abord défini le projet, j'ai échantillonné les différentes espèces de gymnospermes au Jardin Botanique du Parc de la tête d'Or, j'ai réalisé les extractions d'acides nucléiques, j'ai cloné les gènes YABBY par RT-PCR avec des amorces dégénérées (avec l'aide de Nicolas Rode), j'ai obtenu la séquence pleine longueur de ces gènes par RACE (sauf pour un gène étudié par Véronique Boltz), j'ai réalisé les hybridations *in situ* de certains gènes YABBY chez *E. distachya* et *G. biloba*, et j'ai enfin mené de front les reconstructions phylogénétiques des gènes YABBY chez les plantes à graines.

II) Article

Evolution of YABBY gene family in the land plants.

Authors

Cédric Finet, Nicolas Rode, Véronique Boltz, and Charles P. Scutt*

Address

Laboratoire de Reproduction et Développement des Plantes,
Université de Lyon,
(CNRS UMR 5667, INRA, ENS-Lyon, UCB-Lyon),
46 allée d'Italie,
69364 Lyon Cedex 07, France.

***Author for correspondence**

Tel: +33 4 72 72 86 03
Fax: +33 4 72 72 86 00
E-mail: Charlie.Scutt@ens-lyon.fr

Introduction

Six members of the YABBY gene family, which encode putative, plant-specific transcription factors, are present in the *Arabidopsis thaliana* genome: *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER* (*FIL/YAB1*), *YABBY2* (*YAB2*), *YABBY3* (*YAB3*), *YABBY5* (*YAB5*) and *INNER NO OUTER* (*INO*) [1]. The proteins encoded by these genes contain an N-terminal zinc-finger domain, which is hypothesized to function in protein multimerization [2], and a C-terminal "YABBY" domain, related to the DNA-binding domain of High Mobility Group transcription factors, which may therefore also be involved in DNA-binding. YABBY genes are generally expressed in the abaxial (dorsal) tissues of plant lateral organs, such as leaves and floral organs [3]. Though the roles of YABBY genes in plant development are frequently masked by genetic redundancy in single mutants, the overexpression of YABBY genes in plant tissues typically leads to a loss of abaxial-adaxial polarity, producing radially symmetrical lateral organs (leaves etc) made up only of abaxial tissues [4]. Hence, YABBY genes appear to specify abaxial tissue identity in plant lateral organs. Though most YABBY genes are expressed in numerous organ types, *CRC* is expressed only in carpels and nectaries [5]. Strong *crc* mutants show altered gynoecium morphology, including a partial loss of carpel fusion, and the absence of nectaries [6]. Mutations in *crc* interact genetically with mutations in several genes of unrelated families to produce a partial breakdown of adaxial-abaxial polarity in the ovary wall [7]. Similarly, *INO* is expressed only in the outer integument of the ovule. *ino* mutants lack this outer integument [8] : a phenotype which has also been interpreted as a defect in adaxial tissue specification.

A putative ortholog of CRC in the ANA grade angiosperm *Amborella trichopoda* is expressed abaxially in carpel tissues [9], as is *CRC* in *A. thaliana*. Similarly, an *INO* ortholog from the ANA grade angiosperm *Nymphaea alba* is expressed in the outer integument [10], as is *INO* in *A. thaliana,* though also shows some expression in the inner integument and suspensor. It would therefore appear that *CRC* and *INO* have broadly conserved their expression patterns, and possibly their functions, in the *A. thaliana* and in certain basal angiosperm lineages since the last common ancestor of the flowering plants. Evidence of novel expression patterns and functions is, however, available for a *CRC* ortholog from the monocot *Oryza sativa* [11]. This gene, *DROOPING LEAF* (*DL*), is expressed throughout carpel primordia and in leaves. A loss of function in *dl* mutant shows phenotypes in both these organ types that do not appear to relate to a breakdown in abaxial-adaxial polarity. Furthermore, certain YABBY genes in ANA grade angiosperms have been reported to show adaxial expression in lateral organs, in

contrast to the expression of their putative orthologs in *A. thaliana* [12]. Some YABBY genes have also been identified in gymnosperms, though their phylogenetic relationships with angiosperms YABBY genes have not been satisfactorily resolved and their expression patterns have not been reported [13].

The carpel and outer ovule integument are both specific to the angiosperms, though may be homologous to structures in the reproductive axes of gymnosperms. As CRC and INO show specific expression in these two structures, respectively, in angiosperms, we decided to study gymnosperm YABBY genes in more detail to attempt to elucidate the homology of female reproductive structures between distantly related seed plants. We present here the results of this study, which was based on a wide sampling of YABBY genes in the gymnosperms to allow as comprehensive an analysis as possible of the phylogeny of the YABBY family in the seed plants.

Material and methods

Plant materials

Vegetative and reproductive tissues were collected from Lyon Botanic Garden, France. The sampling was representative of the living gymnosperms, including members of Cycadales (*Ceratozamia kuesteriana*, *Cycas thouarsii*), Ginkgoales (*Ginkgo biloba*), Gnetales (*Ephedra distachya*, *Gnetum gnemon*), Coniferales (*Juniperus drupaceae*).

PCR amplification of YABBY cDNAs

Total RNA was isolated from vegetative or reproductive tissues by the method of Chang *et al.* First strand cDNA was synthesized using RevertAid^{TM} M-MuLV Reverse Transcriptase (Fermentas, Burlington, Canada). Partial fragments of YABBY cDNAs were amplified by using the degenerate primers 5'-TTGGATACAGTGACAGTGAAATGYGGNCAITG and 5'-TGCCCAATTTTTTGCTGCTGC. cDNA sequences were completed by 5' and 3' RACE using a MARATHON cDNA amplification kit (Clontech, Palo Alto, USA).

Phylogenetic analyses

Amino acid sequences of the YABBY genes identified were aligned using MUSCLE and manually adjusted. Maximum-likelihood (ML) inferences were performed using PHYML program [14], assuming a WAG+I+Γ_4 model.

Histology

Tissues were fixed in FAA or PFA overnight and then dehydrated through an ethanol series to 100% ethanol. The ethanol was gradually replaced with Histo-Clear II (National Diagnostics). The Histo-clear was gradually replaced with Paraplast X-tra (Fisher Scientific) at 60°C. Embedded specimens were sectioned at 8-10 μm and dried overnight at 37°C. Tissues were hydrated through an ethanol series to H_20 and the slides were stained in 0.1% toluidine blue. Images were captured by using a Leica MZ12 binocular equipped with a Leica DFC320 digital camera.

***In situ* hybridization**

Tissues were fixed in FAA or PFA overnight and then dehydrated through an ethanol series to 100% ethanol. The ethanol was gradually replaced with Histo-Clear II (National Diagnostics). The Histo-clear was gradually replaced with Paraplast X-tra (Fisher Scientific) at 60°C. Embedded specimens were sectioned at 8-10 μm and dried overnight at 37°C. Digoxigenin-labeled antisense and sense RNA probes were prepared from the middle region of *YABBY* genes amplified by PCR. Our prehybridization, hybridization, and posthybridization procedures were based on those of Jackson *et al.* with some modifications. A detailed protocol is available from the authors upon request. Slides were examined and photographed on a Leica MZ12 binocular equipped with a Leica DFC320 digital camera.

Results and Discussion

Members of the YABBY family of transcription factors in *Arabidopsis thaliana* have been shown to be involved in the establishment of adaxial/abaxial polarity in lateral organs. There are six *YABBY* genes in *A. thaliana* : *CRABS CLAW* (*CRC*), *FILAMENTOUS FLOWER* (*FIL*), *YABBY2*, *YABBY3*, *INNER NO OUTER* (*INO*), and *YABBY5*.

We identified YABBY homologues from the non-flowering seed plants *Ceratozamia kuesteriana* (*CK14*), *Cycas thouarsii* (*CT4*, *CT11*, *CTn*), *Ephedra distachya* (*ED1*, *ED2*), *Ginkgo biloba* (*GB1*, *GB3*), *Gnetum gnemon* (*GG1*), and *Juniperus drupaceae* (*JD1*). The sequencing of the genome of the bryophyte *Physcomitrella patens* [15] and the lycophyte *Selaginella moellendorffii* (unpublished genome) revealed the absence of *YABBY* genes in these species, suggesting that this gene family could be specific to the seed plants.

A first phylogeny of the YABBY gene family in the seed plants. At least three different lineages of YABBY genes are present in the extant gymnosperms, with putative orthologues

identified for the *YAB2/YAB5*, *FIL/YAB3* and *INO* clades (Fig 1). No *CRC* orthologue was identified in gymnosperms in spite of extensive and representative sampling, which spanned the whole of gymnosperm phylogeny. According to the topology of our phylogeny, the duplications respectively leading to *YAB2* and *YAB5* and to *FIL* and *YAB3*, occurred after the split between the extant gymnosperms and angiosperms.

We performed in situ hybridization experiments to examine expression patterns of YABBY genes in the gymnosperm *E. distachya*. To better interpret the results of this analysis, we also conducted anatomical analyses of female cone development.

Evolution of *YAB2/YAB5* gene expression in seed plants. Putative *YAB2/YAB5* orthologues were identified from *E. distachya* (*ED2*) and *G. biloba* (*GB2*). Staining of *ED2* mRNA was restricted to the chlamys, the inner bract in Gnetales (Fig 2f, 2g). Unlike *YAB5*, *YAB2* expression has been widely investigated in angiosperms. *YAB2* orthologues have been found to be expressed in the bracts and in the leaves of *A. thaliana* [3], and in the leaves of the basal angiosperm *Amborella trichopoda* [12]. *YAB2* expression has also been reported in the sepals, stamens and carpels of both monocots and eudicots [3, 16, 17]. We may therefore assume that *YAB2/YAB5* genes were involved in bract development of floral organs in the angiosperm lineage.

Signal from the *GB2* probe in female reproductive tissue of *G. biloba* appeared as a light staining in two zones running from the pedicel to both ovules (Fig 2d, 2e). This expression pattern may correspond to provascular tissues. No polarity of expression, nor ovule expression could be observed at early developmental stages. By contrast, strong expression was observed at a late stage of ovule development. Signals in late ovules were restricted to the sclerotesta and nucellus, with no expression detected in the endosperm. Expression of *YAB2* and *YAB5* orthologues in vascular tissues has already been reported in angiosperms. *PROL*, the *YAB5* orthologue from *Antirrhinum majus*, is predominantly expressed in provascular cells during the late stage of apex development [18]. Additionally, a recently published study [17] has shown *OsYABBY1* and *OsYABBY2*, the putative orthologues of *YAB2*, to be expressed in large vascular bundles and in the developing sclerenchyma in rice. These authors noted that *OsYABBY1* was specifically expressed in the mestome sheath in the vascular bundle and in the palea and lemna, which are structures specific to the grass family. The expression of a *YAB2/5* gene in vascular tissues of angiosperms [17] and those of gymnosperms, shown in the present work, suggests an involvement of an ancestral *YAB2/5*

sequence in the control of vascular development in the last common ancestor of the seed plants.

Lastly, *YAB2* expression has been reported to be associated with the establisment of adaxial-abaxial identity in angiosperms [3, 10]. However this does not seem to be the case in the monocots [17]. In the present study, no polarity of expression of *YAB2/YAB5* was observed in gymnosperms. Further expression studies will be needed in order to infer whether *YAB2/YAB5* gene expression was polarized in the ancestor of all seed plants, as has been concluded for class III HD-ZIP genes [19].

Evolution of *FIL/YAB3* gene expression in seed plants. In *E. distachya*, *ED1* is a putative orthologue of *FIL/YAB3*. In young tissues, *ED1* is expressed in the distal region of the ovule. At later developmental stages, *ED1* expression is restricted to the chlamys (Fig 2b, 2c). Expression of *FIL/YAB3* genes has been widely characterized in monocots and eudicots, where they have been found to be expressed in all lateral organ primordia produced from the apical and flower meristems [3, 20-22].

FIL/YAB3 genes have been shown to be expressed in an abaxial manner in eudicots [3, 23]. However, non-polarized expression of these genes has been observed in the lateral organs of monocots [21, 22], with the exception of maize in which the expression of a putative *FIL/YAB3* orthologue was found to be adaxial [20].

Investigating the origin of the outer integument. A putative *INO* orthologue was cloned in *C. thouarsii*, suggesting that the clade INO was present in the last common ancestor of the extant seed plants (Fig 1). Comparison of living angiosperms strongly suggests the last common ancestor to have possessed bitegmic, anatropous ovules, whereas most of gymnosperm ovules are orthotropous and unitegmic (Fig 3). Studies in *A. thaliana* have shown that *ino* mutants failed to develop an outer integument, whereas the inner integument developed normally [8, 24]. Consequently, *ino* mutants have erect, unitegmic ovules that resemble those of gymnosperms. Preliminary characterization of the gene *CT4* indicates this gene is at least expressed in vegetative leaves of *C. thouarsii*. These results suggest a change in the expression and function of INO during the origin of the angiosperms, possibly linked to the origin of the outer integument.

Concluding remarks. The study of YABBY genes in non-flowering plants that we have performed suggests this gene family to have arisen during the early evolution of the seed

plants. Interestingly, the majority of clades of YABBY genes known from model angiosperms (YAB3/FIL, INO, YAB2/5) have also been found in gymnosperms, suggesting the duplications generating these gene clades to have preceded the last common ancestor of the extant seed plants. We have recovered evidence of both conservation and change in YABBY gene function between angiosperms and gymnosperms. For example, we have shown that *YAB2/5* expression in provascular strands is conserved between angiosperms and gymnosperms. This conserved expression pattern reflects an ancestral function in seed plants. By contrast, *INO* appears to have been newly recruited to outer integument development in the angiosperms, as the *INO* ortholog from *Cycas thouarsii* is expressed in vegetative organs.

Acknowledgments

We thank Gilles Dutartre and Jean-Pierre Grienay of Lyon Botanic Garden for assistance with obtaining plants used in this study. We acknowledge core funding from the RDP Laboratory's parent organizations: CNRS, INRA, ENS-Lyon and UCB-Lyon1.

References

1. Golz, J.F., and Hudson, A. (1999). Plant development: YABBYs claw to the fore. Current Biology *9*, R861-R863.
2. Kanaya, E., Watanabe, K., Nakajima, N., Okada, K., and Shimura, Y. (2001). Zinc release from the CH2C6 zinc finger domain of FILAMENTOUS FLOWER protein from Arabidopsis thaliana induces self-assembly. Journal of Biological Chemistry *276*, 7383-7390.
3. Siegfried, K.R., Eshed, Y., Baum, S.F., Otsuga, D., Drews, G.N., and Bowman, J.L. (1999). Members of the YABBY gene family specify abaxial cell fate in Arabidopsis. Development (Cambridge, England) *126*, 4117-4128.
4. Sawa, S., Watanabe, K., Goto, K., Kanaya, E., Morita, E.H., and Okada, K. (1999). FILAMENTOUS FLOWER, a meristem and organ identity gene of Arabidopsis, encodes a protein with a zinc finger and HMG-related domains. Genes & development *13*, 1079-1088.
5. Bowman, J.L., and Smyth, D.R. (1999). CRABS CLAW, a gene that regulates carpel and nectary development in Arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. Development (Cambridge, England) *126*, 2387-2396.
6. Alvarez, J., and Smyth, D.R. (1999). CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. Development (Cambridge, England) *126*, 2377-2386.
7. Eshed, Y., Baum, S.F., and Bowman, J.L. (1999). Distinct mechanisms promote polarity establishment in carpels of Arabidopsis. Cell *99*, 199-209.
8. Villanueva, J.M., Broadhvest, J., Hauser, B.A., Meister, R.J., Schneitz, K., and Gasser, C.S. (1999). INNER NO OUTER regulates abaxial- adaxial patterning in Arabidopsis ovules. Genes & development *13*, 3160-3169.
9. Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C., and Scutt, C.P. (2005). Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel

development since the ancestor of the extant angiosperms. Proceedings of the National Academy of Sciences of the United States of America *102*, 4649-4654.
10. Yamada, T., Ito, M., and Kato, M. (2003). Expression pattern of INNER NO OUTER homologue in Nymphaea (water lily family, Nymphaeaceae). Development genes and evolution *213*, 510-513.
11. Yamaguchi, T., Nagasawa, N., Kawasaki, S., Matsuoka, M., Nagato, Y., and Hirano, H.Y. (2004). The YABBY gene DROOPING LEAF regulates carpel specification and midrib development in Oryza sativa. Plant Cell *16*, 500-509.
12. Yamada, T., Ito, M., and Kato, M. (2004). YABBY2-homologue expression in lateral organs of Amborella trichopoda (Amborellaceae). International Journal of Plant Sciences *165*, 917-924.
13. Lee, J.Y., Baum, S.F., Oh, S.H., Jiang, C.Z., Chen, J.C., and Bowman, J.L. (2005). Recruitment of CRABS CLAW to promote nectary development within the eudicot clade. Development (Cambridge, England) *132*, 5021-5032.
14. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic biology *52*, 696-704.
15. Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science (New York, N.Y *319*, 64-69.
16. Jang, S., Hur, J., Kim, S.J., Han, M.J., Kim, S.R., and An, G. (2004). Ectopic expression of OsYAB1 causes extra stamens and carpels in rice. Plant molecular biology *56*, 133-143.
17. Toriba, T., Harada, K., Takamura, A., Nakamura, H., Ichikawa, H., Suzaki, T., and Hirano, H.Y. (2007). Molecular characterization the YABBY gene family in Oryza sativa and expression analysis of OsYABBY1. Mol Genet Genomics *277*, 457-468.
18. Golz, J.F., Roccaro, M., Kuzoff, R., and Hudson, A. (2004). GRAMINIFOLIA promotes growth and polarity of Antirrhinum leaves. Development (Cambridge, England) *131*, 3661-3670.
19. Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature *428*, 84-88.
20. Juarez, M.T., Twigg, R.W., and Timmermans, M.C. (2004). Specification of adaxial cell fate during maize leaf development. Development (Cambridge, England) *131*, 4533-4544.
21. Zhao, W., Su, H.Y., Song, J., Zhao, X.Y., and Zhang, H.S. (2006). Ectopic expression of TaYAB1, a member of YABBY gene family in wheat, causes the partial abaxialisation of the adaxial epidermises of leaves and arrests the development of shoot apical meristem in Arabidopsis. Plant Science, 364-371.
22. Liu, H.L., Xu, Y.Y., Xu, Z.H., and Chong, K. (2007). A rice YABBY gene, OsYABBY4, preferentially expresses in developing vascular tissue. Development genes and evolution *217*, 629-637.
23. Navarro, C., Efremova, N., Golz, J.F., Rubiera, R., Kuckenberg, M., Castillo, R., Tietz, O., Saedler, H., and Schwarz-Sommer, Z. (2004). Molecular and genetic interactions between STYLOSA and GRAMINIFOLIA in the control of Antirrhinum vegetative and reproductive development. Development (Cambridge, England) *131*, 3649-3659.
24. Baker, S.C., Robinson-Beers, K., Villanueva, J.M., Gaiser, J.C., and Gasser, C.S. (1997). Interactions among genes regulating ovule development in Arabidopsis thaliana. Genetics *145*, 1109-1124.

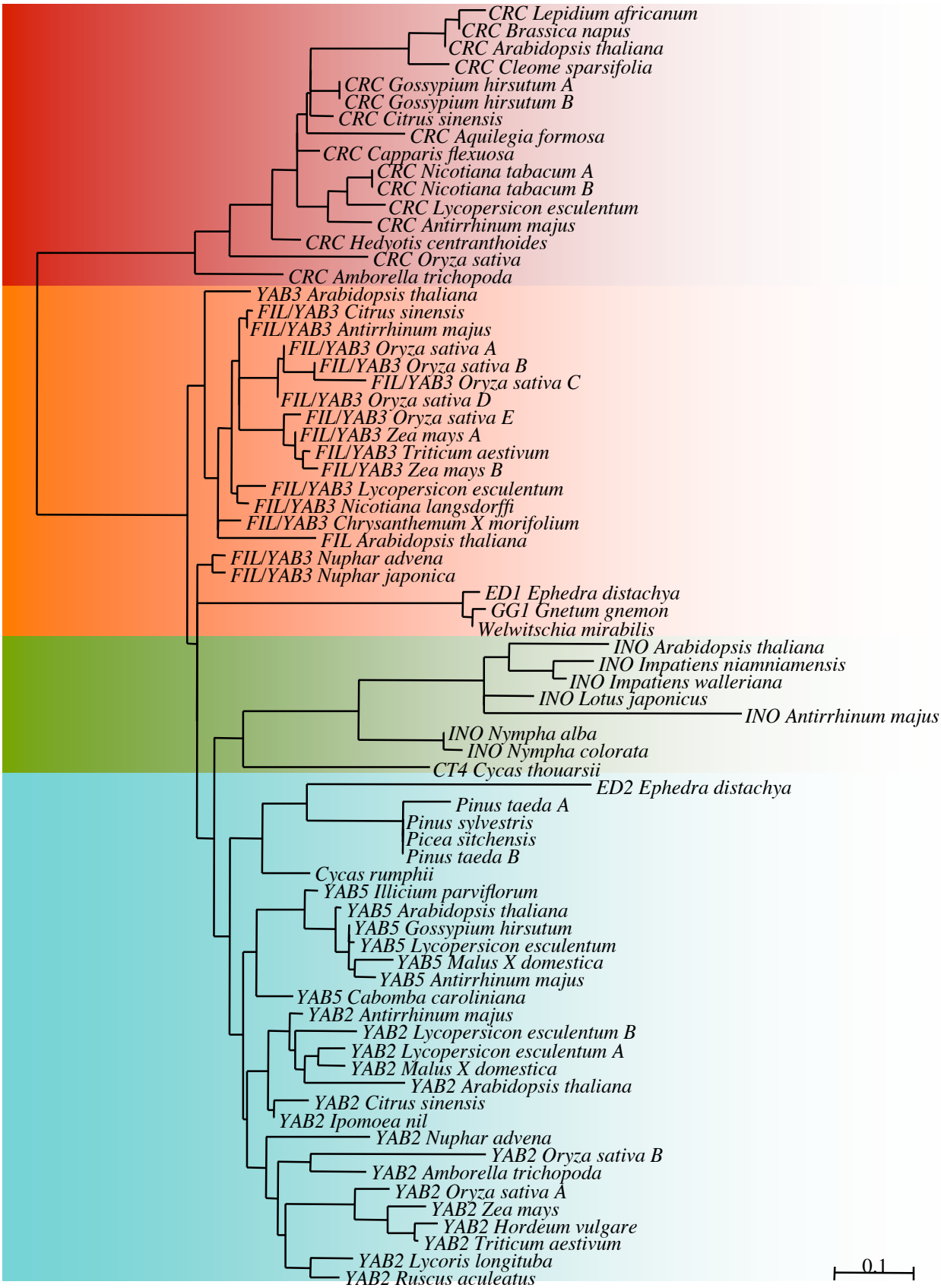


Figure 1. Phylogenetic tree of YABBY genes in the seed plants. ML tree is based on the amino-acid sequences. Partial sequences (*CK14*, *CT11*, *CTn*, *GB1*, *GB3*, *JD1*) were found to bias the results and were therefore excluded from the analysis.

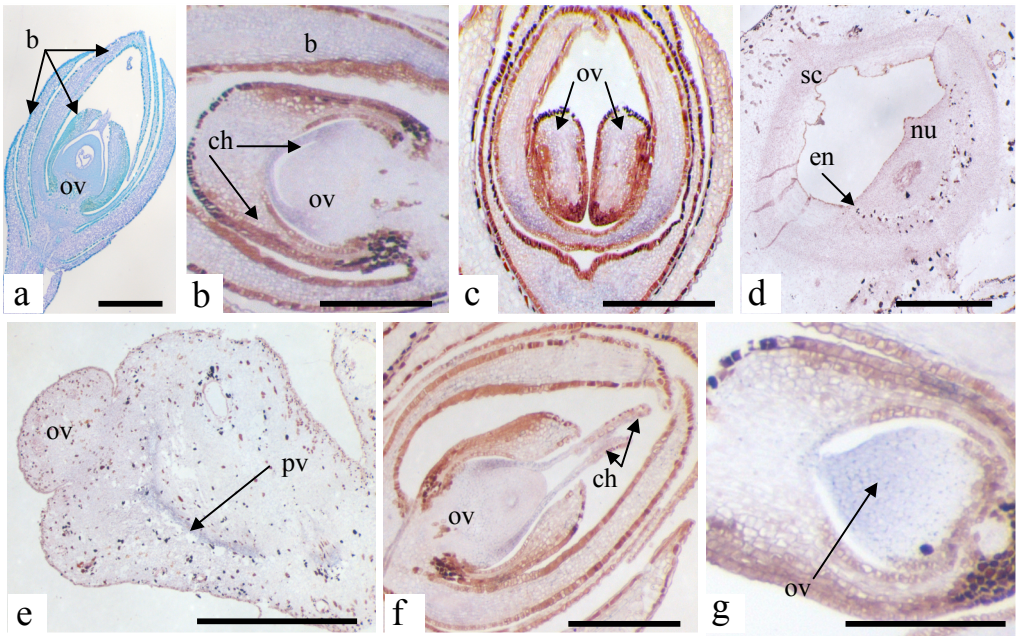


Figure 2. Patterns of RNA *in situ* hybridization of YABBY genes in *E. distachya* and *G. biloba* female cones. *In situ* hybridization signals appear blue or violet. (a) A female cone of *E. distachya* stained in toluidine blue. (b) *In situ* hybridization to a longitudinal section (l.s.) of an *E. distachya* female cone showing *ED1* expression in the chlamys. (c) *In situ* hybridization to a transversal section (t.s.) of an *E. distachya* female cone showing *ED1* expression in ovules and bracts. (d) *In situ* hybridization to a t.s. of a *G. biloba* mature ovule showing *GB2* expression in the sclerotesta and in the nucellus. (e) *In situ* hybridization to a l.s. section of a *G. biloba* female cone showing *GB2* expression in provascular tissues. (f) *In situ* hybridization to a l.s. section of an *E. distachya* female cone showing *ED2* expression in the chlamys. (g) *In situ* hybridization to a l.s. section of an *E. distachya* female cone showing *ED2* expression in the upper part of the ovule. Scale bars represent 0,5mm in a-c, f-g, and 2mm in d-e. b, bract; ch, chlamys; ov, ovule; pv, provascular tissue; sc, sclerotesta.

	gymnosperms	angiosperms
ovules	naked	into the carpel
reproductive axis	unisexual	bisexual
perianth	Ø	present
ovule integument	unique	double
orthotropy	obligatory	rare
campylotropy	Ø	rare
anatropy	Ø	majoritary

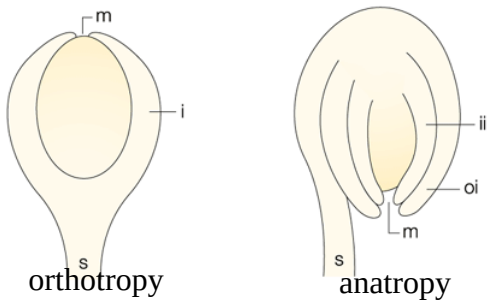


Figure 3. Main features of ovules in angiosperms and gymnosperms.
m, micropyle ; i, inner integument ; oi, outer integument ; s, stem.

III) Perspectives

Les résultats préliminaires de cet article sont les suivants : (i) il devait y avoir au moins trois gènes YABBY chez l'ancêtre commun des plantes à graines actuelles *vs.* cinq gènes chez l'ancêtre commun des angiospermes actuelles, (ii) les gènes YABBY sont caractérisés par une expression polarisée de manière adaxiale ou abaxiale chez les angiospermes alors que ces gènes ne le sont pas chez les gymnospermes. Au regard de ces résultats, on peut se demander si la fonction des gènes YABBY est identique chez l'ensemble des plantes à graines. En d'autres termes, les gènes YABBY ne joueraient peut-être pas un rôle dans la mise en place de la polarité adaxiale/abaxiale des organes latéraux chez les gymnospermes. Il serait intéressant d'étudier la fonction de d'autres orthologues de gènes intervenant dans cette polarité comme par exemple les gènes *KANADI*, *PHABULOSA* ou encore *PHAVOLUTA*.

Cette étude est la seule pour laquelle quelques manipulations restent à faire avant de soumettre le manuscrit. Les protéines YABBYs présentent peu de sites informatifs pour la reconstruction phylogénétique. Leur nombre est tout de même suffisant pour les séquences pleine longueur mais ce n'est pas le cas de toutes les séquences que j'ai cloné durant ma thèse. Ainsi, finir de cloner les extrémités 5' et 3' des ADNc par RACE PCR constituera une des priorités pour ce projet. Cela me permettra d'obtenir un alignement complet de protéines YABBYs et de reconstruire un arbre phylogénétique plus fiable. Toujours dans la même idée, je vais constituer un alignement non plus avec les séquences protéiques mais avec les séquences nucléotidiques afin de tripler le nombre de sites informatifs. Ce jeu de données est actuellement en cours de réalisation. Le jeu de données de gènes YABBY constitué devrait être rendu plus conséquent par l'ajout de données obtenues par le groupe de recherche du professeur Toshihiro Yamada. Cette mise en commun d'informations devrait aboutir à une collaboration scientifique de nos deux groupes.

Concernant l'étude des patrons d'expression, il me reste encore trois gènes à étudier par hybridation *in situ*, ainsi que la caractérisation par RT-PCR semi-quantitative de l'expression de l'orthologue putatif du gène INO chez *C. thouarsii*. Ce résultat est d'autant plus important qu'il permettra peut-être d'inférer une homologie de structure entre le tégument externe de l'ovule des angiospermes et une (des) structure(s) des gymnospermes. L'étude de l'expression de ce gène pourrait également permettre de tester la théorie "Mostly Male". Une expression restreinte de ce gène aux cônes femelles de *C. thouarsii* aurait tendance à confirmer la théorie, alors qu'une expression restreinte aux cônes mâles infirmerait les prédictions de la théorie "Mostly Male".

Chapitre 5 : GÈNES *MIR164* ET ÉVOLUTION DU CARPELLE

I) Introduction et résumé de l'article

Les microARNs matures sont des petits ARNs simple brin d'environ 21 nt, connus pour réguler négativement l'expression de certains gènes cibles avec lesquels ils présentent une complémentarité de séquence (Bartel, 2004). Chez les plantes, la régulation via le micro-ARN fait intervenir soit le clivage de l'ARNm ciblé (Llave et al., 2002; Schwab et al., 2005; Tang et al., 2003), soit l'inhibition de la traduction du gène cible (Brodersen et al., 2008). Les microARNs jouent un rôle central dans le développement des plantes (pour revue voir (Mallory and Vaucheret, 2006)), suggérant que ces molécules ont pu être importantes dans l'évolution des processus développementaux. Des travaux ont cherché à étudier l'évolution de ces microARNs matures soit par recherche du site de liaison dans les orthologues des gènes cibles identifiés chez *A. thaliana* (Axtell and Bartel, 2005), soit par clonage direct des microARNs matures dans des espèces non-modèles (Barakat et al., 2007).

Malgré le développement de ces nouvelles techniques, la séquence du microARN mature ne suffit pas à inférer des hypothèses quant à l'évolution de la famille génique. En effet, il faut bien comprendre que (i) les 21 nucléotides en moyenne du microARN mature sont insuffisants pour la reconstruction phylogénétique, (ii) les différents gènes *MIR* de la famille codent pour le même microARN mature ce qui rend impossible l'étude de l'expression spécifique de chacun de ces gènes (même si l'étude de l'expression globale du microARN mature est possible).

Le papier suivant est novateur dans le sens où il propose une méthode pour cloner les gènes *MIR164* (et non uniquement le *mir164* mature) chez n'importe quelle espèce de plante à fleurs. Le séquençage de ces gènes a permis de mettre en évidence des motifs conservés en dehors des sites miR (21 nt) et miR* (21 nt), rendant ainsi possible la reconstruction phylogénétique de la famille MIR164 chez les angiospermes.

Dans un souci de clarté scientifique, notons que Sophie Jasinski et Aurélie Vialette ont réalisé l'ensemble du travail expérimental et que mon travail a consisté en la reconstruction phylogénétique de la famille MIR164 ainsi que de la réflexion et de la rédaction du manuscrit se rapportant directement à la phylogénie.

II) Article

Full Title

The Evolutionary-Developmental Analysis of Plant MicroRNAs

Running Title

Evo-Devo Analysis of Plant miRNAs

Article Category

Technical Advance

Authors
Sophie Jasinski, Cédric Finet[#], Aurélie C. M. Vialette[#], and Charles P. Scutt*

[#]These two authors contributed equally to this work.

Author Address
Laboratoire de Reproduction et Développement des Plantes,
Université de Lyon, CNRS UMR 5667, INRA, ENS-Lyon, UCB-Lyon,
Ecole Normale Supérieure de Lyon,
46 allée d'Italie,
69364 Lyon Cedex 07, France.

*For correspondence
Tel: +33 4 72 72 86 03
Fax: +33 4 72 72 86 00
E-mail: Charlie.Scutt@ens-lyon.fr

E-mail addresses for co-authors:
Sophie.Jasinski@ens-lyon.fr
Cedric.Finet@ens-lyon.fr
Aurelie.Vialette@ens-lyon.fr

Key Words:
microRNA, evolution, development, phylogeny, angiosperm, MIR164

Total Word count (excluding references and supplementary material): **5143.**
(Breakdown: Summary 248, Introduction 985, Results 2415, Experimental Procedures 1209, Acknowledgements 56, Figure Legends 230.)

Summary

MicroRNAs (miRNAs) control many important aspects of plant development, suggesting these molecules may also have played key roles in the evolution of developmental processes in plants. However, evolutionary-developmental (evo-devo) studies of miRNAs have been held back by technical difficulties in gene identification and phylogenetic analysis. To help to solve these problems, we present here two novel approaches: firstly, we have developed a two-step procedure for the efficient identification of miRNA genes in any plant species and secondly, we have devised a combination of methods that enable the generation of well-supported miRNA gene phylogenies across considerable evolutionary distances. As a test case, we have studied the evolution of the *MIR164* family in the angiosperms. We have identified novel *MIR164* genes in three species occupying key phylogenetic positions: *Amborella trichopoda* and *Cabomba aquatica*, representing the two earliest-diverging lineages of all angiosperms, and *Aquilegia vulgaris*, representing the earliest-diverging lineage of eudicots. Using *MIR164* sequences from these and other species, we have partially reconstructed the evolution of the *MIR164* family in the angiosperms since the last common ancestor of this group. We have furthermore performed expression studies of *MIR164* genes in species occupying key phylogenetic positions and used the results of these analyses to propose hypotheses on the roles *MIR164* genes have played in angiosperm evolution. The techniques we describe may be applied to any miRNA family and should thus enable the discipline of plant evo-devo to investigate the many contributions miRNA genes may have made to the evolution of plant development.

Introduction

Evolutionary-developmental (evo-devo) biology aims to explain how developmental processes evolved (Raff, 2000). Evo-devo analyses frequently begin from an understanding of gene function in model species and a robust species phylogeny in which those models are placed. Orthologous genes are then studied in species occupying key phylogenetic positions to deduce the molecular changes that were responsible for evolutionary events of interest. To date, plant evo-devo has mainly focused on genes encoding developmental regulators such as transcription factors and signal transduction components. However, it is now clear that many plant developmental processes are regulated by microRNAs (miRNAs) (Mallory and Vaucheret, 2006), suggesting these molecules also may have played important roles in the evolution of plant development.

miRNAs are single-stranded RNA molecules of approximately 21 nt that negatively regulate gene expression by hybridizing to complementary target sites in specific messenger RNAs (mRNAs) (Filipowicz *et al.*, 2008). In plants, miRNAs typically show very high similarity to the complement of their target sites in mRNAs, and act by directing the cleavage of these (Jones-Rhoades *et al.*, 2006). miRNAs are transcribed from nuclear genes to generate primary transcripts that are then processed via intermediate forms termed pri- and pre-miRNAs. In these intermediate molecules, a miR site of approximately 21 nt forms a duplex with a partially complementary miR* site. After further processing, the miR-miR* duplex separates and the miR site, corresponding to the mature miRNA, becomes integrated into an RNA-induced silencing complex (RISC) which is then able to target mRNAs in the cytoplasm.

The *MIR164* family in the model angiosperm *Arabidopsis thaliana* consists of three genes, *Ath-MIR164a, b* and *c,* which are potentially capable of targeting seven members (Gustafson *et al.*, 2005) of the *NAC* transcription factor family. Among these targets, the developmental roles of *CUP-SHAPED COTYLEDON1* and *2 (CUC1/2)* (Aida *et al.*, 1997) and *NAC1 (Xie et al., 2000)* are known. The expression domains of *Ath-MIR164a, b and c* partially overlap (Sieber *et al.*, 2007), which may explain the combination of redundant and non-redundant phenotypes observed in *mir164* mutants. For example, only *ath-mir164a* shows an increase in the depth of leaf sinuses (Nikovics *et al.*, 2006), whereas both *ath-mir164a* and *ath-mir164b* show an increase in lateral root branching (Guo *et al.*, 2005). Other *mir164* mutants show additive phenotypes. For example, *ath-miR164c* shows a slight defect in carpel fusion (Baker *et al.*, 2005), whereas the *ath-mir164abc* triple mutant shows complete carpel separation (Sieber *et al.*, 2007). The distinct functions of *MIR164* family members in *Arabidopsis thaliana* illustrate the need for evo-devo studies of miRNA genes in plants. The processes of gene duplication, gene loss, sub-functionalization and neo-functionalization, long known to have shaped families of protein-coding genes (Ohno, 1970), have clearly also shaped miRNA families, with potentially important consequences for the evolution of plant development.

Most of the plant miRNA genes currently known were identified by the *in silico* searching of partial or complete genome sequences (Bonnet *et al.*, 2004; Jones-Rhoades and Bartel, 2004; Adai *et al.*, 2005; Li and Zhang, 2005; Li *et al.*, 2005). To study the evolutionary importance of miRNAs, methods are required for the identification of miRNA genes in plants occupying key phylogenetic positions whose genomes have not yet been sequenced. The high-throughput sequencing of small RNAs (Barakat *et al.*, 2007; Buhtz *et al.*, 2008) can be used to

identify mature miRNAs in any species. However, the sequences obtained by this method are too short for phylogenetic analysis, and are therefore of limited value to evo-devo studies. Even in cases where miRNA gene or precursor sequences are available, phylogenetic analyses have, to date, given rather unsatisfactory results, in part due to the generally low level of sequence conservation between homologous miRNA genes, other than in their miR and miR* sites (Mica *et al.*, 2006; Zhang *et al.*, 2006a).

To help resolve these problems, we present here methods for the efficient evo-devo analysis of miRNA genes in plants. Firstly, we have devised a two-step procedure for the identification of any miRNA gene in any plant species. This method is specifically applicable to plants due to the generally high conservation of homologous miR sites throughout the plant kingdom. Secondly, we have devised a combination of *in silico* methods that enable the generation of well-supported phylogenies of miRNA genes over considerable evolutionary distances. As a test case, we have analyzed the evolution of the *MIR164* family throughout the extant angiosperms, or flowering plants. To analyze the base of the angiosperm tree, we incorporated in our studies two representatives of the ANA (Amborellales-Nymphaeales-Austrobaileyales) grade, which represents the three earliest-diverging lineages of living angiosperms (Stevens, 2001). The ANA grade models used in this study were *Amborella trichopoda*, a shrub endemic to the rainforests of New Caledonia and the only known member of Amborellales, and *Cabomba aquatica*, a small aquatic plant native to Brazil and a convenient representative of Nymphaeales. We also analyzed the *MIR164* family in *Aquilegia vulgaris,* which occupies a key phylogenetic position as a member of the most basally-diverging eudicot order, Ranunculales (Stevens, 2001). Our analysis of *MIR164* phylogeny has enabled us to draw firm conclusions on the structure of the *MIR164* family in the last common ancestor of the extant angiosperms and to partially reconstruct the sequence of gene duplications and losses that have led to the present-day *MIR164* families in *Arabidopsis thaliana* and other model species. We have performed gene-specific expression analyses of *MIR164* sequences and superimposed the data obtained from these onto our *MIR164* phylogeny to identify conservation and change in expression patterns during angiosperm evolution. We use the results of our analyses to propose testable hypotheses on the roles of *MIR164* genes in the evolution of the angiosperms. Our study represents a paradigm case for the evolutionary-developmental analysis of miRNA genes. The methods we have developed should be applicable to any miRNA gene family in any plant group and should thus enable studies of the roles that these important regulatory genes have played in plant evolution.

Results

A two-step procedure for the identification of miRNA genes in any plant species.

We have developed a two-step procedure for the identification of miRNA genes of known families in any plant species. The first step of this procedure consists of the PCR suppression-amplification (Broude *et al.*, 2001) of miRNA gene fragments from adaptor-ligated plant genomic DNA using a miRNA-specific PCR primer in combination with an adaptor primer that is common to all DNA molecules. Such amplifications typically yield mixtures of PCR products containing a proportion of miRNA gene fragments. In the second step of our procedure, a database is generated from sequences obtained by PCR suppression-amplification, and then searched for the miR* sites present in miRNA gene fragments. To quantify the efficiency of the PCR suppression-amplification step of our procedure, we used this to re-identify known *MIR164* genes from *Arabidopsis thaliana* and *Zea mays*. PCR products generated from genomic DNA of these species were ligated into plasmid vectors, cloned in *E. coli*, and screened by colony hybridization using *MIR164* gene-specific radiolabeled probes. All six *MIR164* probes used, *Ath-MIR164a, b* and *c*, and *Zma-MIR164a, b* and *d* (Griffiths-Jones *et al.*, 2008), revealed positively hybridizing bacterial colonies at frequencies of between 0.1 and 8 % (Suppl. Tab. 1).

Following the success of the above test, we employed our full, two-step procedure to identify novel *MIR164* genes from *Amborella trichopoda*, *Cabomba aquatica* and *Aquilegia vulgaris*. The PCR products inserted into approximately 220 recombinant plasmids, generated from each of these species as described above, were sequenced from one extremity. A DNA databases was compiled from sequences of each species and searched using an optimized BLAST protocol. This procedure resulted in the identification of three putative *MIR164* genes from each of *Aquilegia vulgaris* (*Avu-MIR164a-c*) and *Cabomba aquatica* (*Caq-MIR164a-c*), and one such gene from *Amborella trichopoda (Atr-MIR164a)*. The putative miRNA genes identified were present in populations of sequenced plasmids at frequencies of between 0.4 and 2 % (Suppl. Tab. 1), in broad agreement with our earlier trial runs using *Arabidopsis thaliana* and *Zea mays*. Further rounds of genomic PCR were then used to complete the sequences of the putative *MIR164* genes identified.

The predicted secondary structures and minimal free energy indices (MFEI) of the novel sequences identified (Fig. 1) are in good agreement with their designation as *MIR164* genes

(Ambros *et al.*, 2003). PCR amplifications are known to be particularly sensitive to mismatches involving one or both of the two 3'-terminal bases of PCR primers. Accordingly, the PCR amplifications described here were performed with two versions of the *miR164* primer, one of which lacked the last two bases of the *miR164* consensus sequence. The genes *Avu-MIR164a*, *Caq-MIR164b* and *Caq-MIR164c* were obtained only from amplifications using the shortened PCR primer and were subsequently found to differ from the consensus *miR164* sequence at their 3'-extremities (Fig. 1), indicating the usefulness of this approach. The two-step procedure described here should prove widely applicable for the identification of miRNA genes in any plant species, thereby fulfilling the first essential prerequisite for the evo-devo analysis of plant miRNAs.

A procedure for the phylogenetic analysis of miRNA genes over long evolutionary distances.

To maximize the resolution of phylogenetic analyses of miRNA genes, we made an initial alignment of *MIR164* family sequences using a computer-based procedure, and then improved the alignment of conserved blocks of nucleotides by extensive manual adjustment. Various models for nucleotide substitution during evolution were then tested for their goodness-of-fit to the finished dataset. On the basis of these tests, we made an initial Maximum Likelihood analysis of the *MIR164* dataset (Suppl. Fig. 1) assuming a GTR+Γ_4 evolutionary model. This analysis generated a phylogeny in which one major clade was well-resolved, though sequences external to this clade were not (Suppl. Fig. 2a). We designated the well-resolved clade in this phylogeny as the B-clade, after the *Arabidopsis thaliana* gene *Ath-MIR164b* within it. We then reanalyzed the same dataset, assuming a slightly different evolutionary model, GTR+I+Γ_4. This procedure permitted the resolution of the sequences external to the previously identified B-clade, though did not correctly resolve the phylogeny of B-clade sequences (Suppl. Fig. 2b). The difficulty we encountered in resolving the entire gene phylogeny in one step very probably resulted from long-branch attraction between the two major clades of sequences identified in the two analyses performed. To overcome this problem, we assembled the two well-resolved sub-trees into a composite phylogeny of *MIR164* genes (Fig. 2a). We designated two well-supported clades of the second sub-tree of this composite phylogeny as the A- and C-clades, after the *Arabidopsis thaliana* genes *Ath-MIR164a* and *c*, respectively within them. We designated a further major clade of this sub-tree, which contained no genes from *Arabidopsis thaliana*, as the D-clade. Analyses performed using other well-fitting evolutionary models produced phylogenies (data not shown) that were highly congruent with that in Fig. 2a.

Comparison of our phylogeny of *MIR164* genes (Fig. 2a) with the current view of angiosperm phylogeny (Fig. 2b) has allowed a number of conclusions to be drawn on the evolution of the *MIR164* family in the angiosperms. One of the clearest features of our analyses was the very well-supported basal separation of the B-clade from all other clades of *MIR164* genes. B-clade genes are present in all the major angiosperm groups analyzed, including ANA grade angiosperms (*Amborella trichopoda, Cabomba aquatica*), Poaceae (*Oryza sativa, Sorghum bicolor, Triticum aestivum, Zea mays*), basal eudicots (*Aquilegia vulgaris*) and core eudicots (*Arabidopsis thaliana*, *Populus trichocarpa, Vitis vinifera).* The internal structure of the B-clade broadly recapitulates the consensus view of angiosperm phylogeny (Fig. 2b). Accordingly, genes from ANA grade angiosperms occupy basal positions in the B-clade, while most of the remaining B-clade genes divide perfectly into two sub-clades, respectively containing only monocot and core eudicot sequences. These data clearly suggest a B-lineage of *MIR164* genes to have diverged from other *MIR164* lineages prior to the last common ancestor of the extant angiosperms. The B-lineage has become duplicated several times during angiosperm evolution to generate two or more B-clade genes in monocots and in the eudicot, *Populus trichocarpa*.

The *MIR164* A- and C-clades (Fig. 2a) occur in sister positions and both contain genes from several eudicots, though not from monocots or ANA grade angiosperms. The monocot gene, *Osa-MIR164c,* occurs with good bootstrap support is a sister position to the combined A/C clade. Taken together, these data suggest the ancestral genes of the A- and C-clades to have been generated by a duplication event in the eudicot lineage, after its divergence from that of the monocots. The timing of this A/C duplication, relative to the separation of the basal and core eudicot lineages, could not be determined due to low bootstrap support for nodes within the A- and C-clades. The D-clade occurs in a paraphyletic position to the ancestral A/C lineage (Fig. 2a). This clade contains genes from an ANA grade angiosperm (*Cabomba aquatica*), monocots and eudicots, suggesting it to have separated from the A/C-lineage before the radiation of the extant angiosperms. The presence of a D-clade gene in *Populus trichocarpa* (eurosids II), but not in *Arabidopsis thaliana* (eurosids I), strongly suggests the D-clade to have been lost from the *Arabidopsis thaliana* lineage since the separation of eurosids I and II.

Comparison of phylogenetically informative taxa permits the deduction of ancestral miRNA gene expression patterns in the angiosperms.

We used semi-quantitative RT-PCR to analyze the expression patterns of *MIR164* genes in *Arabidopsis thaliana, Cabomba aquatica, Aquilegia vulgaris* and *Amborella trichopoda* (Fig. 3). B-clade genes of the *MIR164* family were expressed at remarkably similar levels in the leaves, flowers and roots of all these species (Fig. 3a-d), except for a somewhat lower level of expression in the roots of *Amborella trichopoda* (Fig. 3d). B-clade gene expression was easily revealed in all species tested, suggesting genes of this clade to be more highly expressed than those of other clades. Taken together, these data suggest the conservation of strong B-clade gene expression in leaves, flowers and roots of the plant lineages tested, since the initial the radiation of the angiosperms.

The A- and C-clade genes from *Arabidopsis thaliana, Ath-MIR164a* and *c,* and an A-clade gene from *Aquilegia vulgaris, Avu-MIR164b,* were expressed at comparable levels in leaf and flower tissues, though much less expressed in roots (Fig. 3a and b). The broad similarity in the expression patterns of A- and C-clade genes suggests these to have been conserved since the gene duplication event that separated their lineages, which our phylogenetic studies indicate to have occurred in the eudicots (Fig. 2a).

A D-clade gene from *Cabomba aquatica, Caq-MIR164b,* was expressed in both leaf and flower tissues (Fig. 3c), whereas that from *Aquilegia vulgaris, Avu-MIR164a,* was measurably expressed only in leaves (Fig. 3d). It is not possible, in the absence of D-clade expression data from other basally-diverging lineages, to conclude which, if either, of these patterns might represent the ancestral condition in the angiosperms. A pronounced foliar dimorphism is present in *Cabomba aquatica*, which possesses submerged leaves that are very highly dissected, and floating leaves with entire margins. *Caq-MIR164b* expression was detected in the submerged leaves, but not in the floating leaves, of *Cabomba aquatica* (Fig. 3c). Similarly, the B-clade gene, *Caq-MIR164a,* was more highly expressed in submerged than in floating leaves (Fig. 3c). These preliminary data, together with the known effects of *mir164* on leaf morphology in *Arabidopsis thaliana* (Nikovics *et al.*, 2006), suggest a possible role for *mir164* in the establishment of leaf dimorphism in *Cabomba aquatica.*

A second D-clade gene from *Cabomba aquatica, Caq-MIR164c,* showed no measurable expression in our analyses (Fig. 3c). Demonstrable expression has been proposed as a prerequisite for the identification of novel miRNA genes (Ambros *et al.*, 2003). However, *Caq-MIR164c* may be considered as a *MIR164* gene as it occupies a well-supported position

in phylogenetic reconstructions of the *MIR164* family (Fig. 2a). This gene may have a very specific expression profile, which was not revealed in the analyses presented here, or may alternatively represent a recently generated pseudogene which is no longer expressed, but whose sequence has not yet become substantially degraded.

Discussion

Novel methods have permitted the reconstruction of *MIR164* evolution since the last common ancestor of the living flowering plants.

We have devised a combination of methods for the identification and phylogenetic analysis of miRNA genes in plants, which should facilitate studies of the roles these genes have played in plant evolution. We have demonstrated the efficiency of our novel procedures by partially reconstructing the evolution of the *MIR164* family since last common ancestor of the extant angiosperms, which is estimated to have lived some 160 million years ago (Davies *et al.*, 2004). According to our analyses, three lineages of *MIR164* genes, the A/C-, B- and D-lineages, would have been present in the last common ancestor of the angiosperms (Fig. 4). Our studies have partially reconstructed the sequence of gene duplications and losses that have led to the *MIR164* families of present-day model plants. Accordingly, descendents of all three ancestral *MIR164* genes are present in the monocot *Oryza sativa*, and in the core eudicots *Populus trichocarpa* and *Vitis vinifera*. However, only the A/C- and B-lineages are present in the core eudicot *Arabidopsis thaliana*, the D-clade having been lost from this lineage subsequent to the separation of eurosids I and II. The A/C-lineage is represented by a single gene in the monocot *Oryza sativa*, but by separate A- and C-clade genes in several widely-diverged eudicots. Hence, the A- and C- lineages of *MIR164* genes can be hypothesized to have been generated by a gene duplication event that occurred at an early stage in the evolution of the eudicots (Fig. 4).

To study conservation and change in *MIR164* gene expression during angiosperm evolution, we performed semi-quantitative RT-PCR analyses of *MIR164* sequences in several species occupying key phylogenetic positions (Fig. 3). The results of these studies, superimposed of our phylogeny of *MIR164* genes (Fig. 2a), revealed the probable conservation of B-clade *MIR164* gene expression, at a high level in leaves, flowers and roots, dating back to the last common ancestor of the extant angiosperms (Fig. 4). Our analyses also revealed a higher level of A- and C-clade gene expression in leaves and flowers than in roots, which may reflect the ancestral condition of the A/C-lineage in early eudicots (Fig. 4). By contrast to the members

of other *MIR164* clades, our analyses showed D-clade gene expression to be more variable in the species studied, and did not allow any general conclusions to be drawn.

What roles have *MIR164* genes played in the evolution of the angiosperms?

In *Arabidopsis thaliana,* the balance of expression between *Ath-MIR164a* and *CUC2* is known to control the depth of leaf sinuses (Nikovics *et al.,* 2006). In this study, we have identified a marked difference in the levels of *miR164* between the two leaf forms of *Cabomba aquatica,* which also differ considerably in their degree of dissection. Hence, the preliminary results obtained here may indicate roles for *miR164* and its *NAC* family targets in leaf dissection in an ANA grade angiosperm. Leaf lobing and dissection are highly variable traits in the angiosperms (Bharathan *et al.,* 2002). The possible involvement of the *MIR164/NAC* family expression balance in the control of leaf development in species as distantly related as *Arabidopsis thaliana* and *Cabomba aquatica* suggests this mechanism may have been independently recruited many times throughout the evolution of the angiosperms. Further studies will be needed to confirm the roles of *MIR164* and *NAC* genes in *Cabomba aquatica* leaf development, and might also then be extended to other species showing parallel evolution of dissected leaf morphology.

The major defining feature of the angiosperms is the carpel, or specialized female reproductive organ that encloses the ovules. In gymnosperms, which form the sister group to the angiosperms, ovules typically occur as naked structures, borne in the axils of open sporophylls. The process of carpel fusion is known to be controlled in *Arabidopsis thaliana* by the relative expression levels of *MIR164* and *CUC2* (Baker *et al.,* 2005; Nikovics *et al.,* 2006; Sieber *et al.,* 2007). Further studies may now be undertaken to determine whether a similar mechanism, involving *MIR164* genes identified in the present work, might be responsible for the closure of the simple carpels present in *Amborella trichopoda* and *Cabomba aquatica*. Evidence of such a mechanism in these ANA grade angiosperms would support the hypothesis that a change in the expression balance of *MIR164* and *NAC* family genes in, for example, the ovule-bearing sporophylls of a gymnosperm-like ancestor, could have generated the closed carpel, and hence the first flowering plants (Frohlich and Chase, 2007).

Experimental Procedures

Plant material

Plants and tissues of *Amborella trichopoda* and *Aquilegia vulgaris* were field-collected from Col d'Amieu (New Caledonia) and from Thoirette (Jura, France), respectively. Seed of the Columbia-0 ecotype of *Arabidopsis thaliana* (accession no. N1092) was obtained from the Nottingham Arabidopsis Stock Centre. Plants of the above species were grown in peat-based compost under greenhouse conditions. Plants of *Cabomba aquatica* were obtained from Anthias S.A. (Les Chères, France) and grown in a freshwater aquarium.

PCR suppression-amplification of miRNA gene fragments.

Genomic DNA was extracted from leaf tissues of *Amborella trichopoda, Aquilegia vulgaris, Cabomba aquatica* and *Arabidopsis thaliana* using a NUCLEON PHYTOPURE kit (Amersham), incorporating an RNAase digestion step. Genomic DNA of *Zea mays* inbred line A188 (Gerdes and Tracy, 1993) was kindly provided by Dr Peter Rogowsky (RDP, Lyon). To amplify miRNA gene sequences, PCR-suppression amplifications were performed, based on the method of Broude *et al.* (2001). Aliquots of genomic DNA (10 µg) from all species studied were digested with *Dra*I or *Hae*III. Double-stranded DNA adaptors (Broude *et al.*, 2001) were ligated to the resulting cleaved DNA fragments, and unincorporated adaptors were then removed using NUCLEOSPIN PCR clean-up columns (Machrey-Nagel). PCR amplifications were performed from aliquots (25 ng) of adaptor-ligated DNA, using a 21 nt primer corresponding to the consensus *miR164* sequence from *Arabidiopsis thaliana* (Fig. 1) and the A-adaptor primer (Broude *et al.*, 2001). Amplifications were also performed using a 19-mer *miR164* primer, lacking the two bases at the 3'-extremity of the consensus *miR164* sequence. PCR amplifications were performed using 0.02 units/µL of EX-TAQ Polymerase (Takara), 0.2 µM of each primer, 0.2 mM dNTPs and reaction buffer (1x), as supplied by the manufacturer and continued for 30 thermal cycles, each consisting of a denaturation step of 94°C for 30 s, a thermal gradient annealing step of 45-65°C for 30 s, and an elongation step of 72°C for 90 s. Aliquots of the PCR products generated were analyzed on 1% agarose gels. DNA samples for further study were selected from amplifications performed at the highest annealing temperature that yielded products in each case. Aliquots (1 µl) of the DNA samples selected were ligated into the *pCRII* vector (Invitrogen) and the recombinant plasmids generated were used to transform *E. coli* DH5α competent cells and plated on ampicillin-containing media.

To estimate the proportion of *MIR164* gene fragments in PCR products amplified from *Arabidopsis thaliana* and *Zea mays* genomic DNA, 520 and 1040 ampicillin-resistant colonies

generated respectively from these two species were transferred onto HYBOND-N (Amersham) membranes for colony hybridizations using radio-labeled, gene-specific probes of *Ath-MIR164a, b* and *c* (for *Arabidopsis thaliana* genes), or *Zma-MIR164a, b* and *d* (for *Zea mays* genes). To identify novel *MIR164* genes from *Amborella trichopoda, Cabomba aquatica* and *Aquilegia vulgaris,* approximately 220 antibiotic-resistant bacterial colonies from each species, generated as described above, were sent to a commercial service provider for plasmid purification and DNA sequencing using the M13 forward sequencing primer. PCR products derived from DNA cleaved by *Dra*I and *Hae*III, and from amplifications involving full-length and shortened *miR164* primers, were selected for sequencing in approximately equal proportions.

Database construction and BLAST searching for miR* sites.

DNA sequences of PCR products, generated as described above, were imported into CONTIG EXPESS (Invitrogen). Cloning vector regions were removed, and the resulting sequences were assembled into contigs. For contigs containing a *miR164* primer at the 5'-end, only the longest contributing sequence was retained for further analysis, whereas for contigs beginning with an adaptor primer, the whole contig was saved, reversed and complemented. The resulting sequences from each plant species were exported as a FASTA file and aligned over their terminal *miR164* primer sequences, where present, using MUSCLE (Edgar, 2004), which allowed these primer to be removed easily from the aligned sequences. The resulting primer-less sequences were then compiled into stand-alone DNA databases using FORMATDB. The DNA databases generated in this way were searched by BLAST (Altschul *et al.*, 1997) for internal miR* sites using the mature *miR164* consensus sequence (Fig. 1). BLAST options were used to specify searching of the bottom strand only (S2), with a word-size of 6 (W6), no filtering (FF), giving a score of 4 for an exact match (r4), and a penalty of 5 for each mismatch (q-5). BLAST hits were used to search the available plant nucleotide databases to eliminate any false positives that were homologous to known genes. The remaining sequences, representing putative *MIR164* genes, were retained for further study.

PCR amplification across miR sites.

To determine the sequences of mature miR sites present in the putative *MIR164* genes identified, and of the regions immediately adjacent to these, a second round of PCR suppression-amplification was undertaken. These amplifications were performed on further aliquots of adaptor-ligated genomic DNA by the sequential use of two miRNA gene-specific

primers derived from the novel sequences obtained, in combination with the A-adaptor primer (Broude *et al.*, 2001). The resulting PCR products were cloned in plasmid vectors and sequenced, and the novel data obtained were used to complete the sequences of the putative *MIR164* genes identified.

RNA secondary structure prediction and stability calculations

Secondary structures of putative *pre-miR164* molecules were predicted in MFOLD 3.2 (Zuker, 2003) using default parameters. Minimal free energy indices (MFEI) for the structures generated were calculated as described by Zhang *et al.* (2006b).

Phylogenetic analyses of miRNA genes

Sequence alignments were performed using MUSCLE (Edgar, 2004), followed by manual adjustment of the region between miR and miR* sites to generated the alignment shown in Suppl. Fig. 1. MODEL-TEST v.3.07 (Posada and Crandall, 2004) was then used to identify the best-fitting model of DNA substitution for use in phylogenetic analyses by calculating hierarchical likelihood ratios. On the basis of these tests, TrNef+Γ (Tamura and Nei, 1993), a time reversible model with equal frequency substitution and a gamma shape distribution of shape parameter (α) equal to 0.4748, was chosen. A subsequent analysis used the similar model, TrNef+I+Γ, which additionally adjusted for the number of invariable sites (I). Maximum Likelihood (ML) phylogenetic analyses were performed using both TREEFINDER V.-May 2006 (Jobb *et al.*, 2004) and PAUP* V.4.0d90 (Swofford, 2003), with bootstrap analyses of 1000 replicates. ML trees generated in initial studies were used as further input for subsequent ML runs to search for trees with better -ln likelihoods. Other evolutionary models assessed using ModelTest v.3.07, including GTR (Rodriguez *et al.*, 1990) and HKY (Hasegawa *et al.*, 1985), were also used in ML analyses to assess the congruency of the topologies obtained.

Semi-quantitative Reverse Transcriptase-PCR analyses.

Plant tissues were ground under liquid nitrogen, together with 0.1 g of polyvinylpyrrolidone (mw 40 kDa) per gram of tissue, and RNA was extracted from these tissue powders using an RNA-EASY Kit (Qiagen). Samples (0.5 μg) of total RNA were then treated with RNase-free DNAse (Ambion), and each divided into two aliquots. These aliquots were processed for first-strand cDNA synthesis, respectively with and without the addition of REVERT-AID M-

MuLV Reverse Transcriptase (Fermentas), using oligo-dT primers and other reaction components as described by the manufacturer. Appropriate volumes of diluted cDNA samples, and equal volumes of similarly diluted negative control samples (not treated with Reverse Transcriptase), were used as templates in PCR amplifications using *MIR164* gene-specific and *GAPDH* primers. Full details of the primer sequences and thermal cycle conditions used are given in Suppl. Tab. 3.

Acknowledgements

We thank Patrick Laufs for helpful discussions and for supplying *Arabidopsis thaliana MIR164* gene probes, Guy Perrière for bioinformatics advice, and the technical and secretarial staff of The RDP Laboratory. SJ was supported by an ATER post of the ENS-Lyon. CF and AV receive doctoral thesis funding from the French Ministry of Research and Higher Education.

References

Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V., and Sundaresan, V. (2005). Computational prediction of miRNAs in *Arabidopsis thaliana*. Genome Res. **15,** 78-91.

Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in *Arabidopsis*: An analysis of the *cup-shaped cotyledon* mutant. Plant Cell **9,** 841-857.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25,** 3389-3402.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X.M., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003). A uniform system for microRNA annotation. RNA- A Publication of the RNA Society **9,** 277-279.

Baker, C.C., Sieber, P., Wellmer, F., and Meyerowitz, E.M. (2005). The early extra petals1 mutant uncovers a role for MicroRNA miR164c in regulating petal number in *Arabidopsis*. Curr. Biol. **15,** 303-315.

Barakat, A., Wall, K., Leebens-Mack, J., Wang, Y.J., Carlson, J.E., and de Pamphilis, C.W. (2007). Large-scale identification of microRNAs from a basal eudicot (*Eschscholzia californica*) and conservation in flowering plants. Plant J. **51,** 991-1003.

Bharathan, G., Goliber, T.E., Moore, C., Kessler, S., Pham, T., and Sinha, N.R. (2002). Homologies in leaf form inferred from *KNOXI* gene expression during development. Science **296,** 1858-1860.

Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. (2004). Detection of 91 potential in plant conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. Proc. Natl. Acad. Sci. U. S. A. **101,** 11511-11516.

Broude, N.E., Zhang, L.G., Woodward, K., Englert, D., and Cantor, C.R. (2001). Multiplex allele-specific target amplification based on PCR suppression. Proc. Natl. Acad. Sci. U. S. A. **98,** 206-211.

Buhtz, A., Springer, F., Chappell, L., Baulcombe, D.C., and Kehr, J. (2008). Identification and characterization of small RNAs from the phloem of *Brassica napus*. Plant J. **53,** 739-749.

Davies, T.J., Barraclough, T.G., Chase, M.W., Soltis, P.S., Soltis, D.E., and Savolainen, V. (2004). Darwin's abominable mystery: Insights from a supertree of the angiosperms. Proc. Natl. Acad. Sci. U. S. A. **101,** 1904-1909.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. Bmc Bioinformatics **5,** 1-19.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nature Reviews Genetics **9,** 102-114.

Frohlich, M.W., and Chase, M.W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. Nature **450,** 1184-1189.

Gerdes, J.T., and Tracy, W.F. (1993). Pedigree diversity within the Lancaster Surecrop heterotic group of maize. Crop Science **33,** 334-337.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res. **36,** D154-D158.

Guo, H.S., Xie, Q., Fei, J.F., and Chua, N.H. (2005). MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell **17,** 1376-1386.

Gustafson, A.M., Allen, E., Givan, S., Smith, D., Carrington, J.C., and Kasschau, K.D. (2005). ASRP: the Arabidopsis Small RNA Project Database. Nucleic Acids Res. **33,** D637-D640.

Hasegawa, M., Kishino, H., and Yano, T.A. (1985). Dating of the Human Ape Splitting by a Molecular Clock of Mitochondrial-DNA. J. Mol. Evol. **22,** 160-174.

Jobb, G., von Haeseler, A., and Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. Bmc Evolutionary Biology **4**.

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant MicroRNAs and their targets, including a stress-induced miRNA. Mol. Cell **14,** 787-799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. Annual Review of Plant Biology **57,** 19-53.

Li, X., and Zhang, Y.Z. (2005). Computational detection of microRNAs targeting transcription factor genes in *Arabidopsis thaliana*. Computational Biology and Chemistry **29,** 360-367.

Li, Y., Li, W., and Jin, Y.X. (2005). Computational identification of novel family members of MicroRNA genes in *Arabidopsis thaliana* and *Oryza sativa*. Acta Biochimica Et Biophysica Sinica **37,** 75-87.

Mallory, A.C., and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. Nat. Genet. **38,** S31-S36.

Mica, E., Gianfranceschi, L., and Pe, M.E. (2006). Characterization of five microRNA families in maize. Journal of Experimental Botany **57,** 2601-2612.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the *MIR164A* and *CUC2* genes controls leaf margin serration in *Arabidopsis*. Plant Cell **18,** 2929-2945.

Ohno, S. (1970). Evolution by Gene Duplication. (Berlin: Springer-Verlag).

Posada, D., and Crandall, K.A. (2004). Modeltest: testing the model of DNA substitution. Bioinformatics **14,** 817- 818.

Raff, R.A. (2000). Evo-devo: the evolution of a new discipline. Nature Reviews Genetics **1,** 74-79.

Rodriguez, F., Oliver, J.L., Marin, A., and Medina, J.R. (1990). The General Stochastic-Model of Nucleotide Substitution. J. Theor. Biol. **142,** 485-501.

Sieber, P., Wellmer, F., Gheyselinck, J., Riechmann, J.L., and Meyerowitz, E.M. (2007). Redundancy and specialization among plant microRNAs: role of the *MIR164* family in developmental robustness. Development **134,** 1051-1060.

Stevens, P.F. (2001). Angiosperm Phylogeny Website, Version 7. http://www.mobot.org/MOBOTresearch/APweb/.

Swofford, D.L. (2003). PAUP* V4.Od90. Phylogentic analysis using parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland MA.

Tamura, K., and Nei, M. (1993). Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-DNA in Humans and Chimpanzees. Mol. Biol. Evol. **10,** 512-526.

Xie, Q., Frugis, G., Colgan, D., and Chua, N.H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev. **14,** 3024-3036.

Zhang, B.H., Pan, X.P., Cannon, C.H., Cobb, G.P., and Anderson, T.A. (2006a). Conservation and divergence of plant microRNA genes. Plant J. **46,** 243-259.

Zhang, B.H., Pan, X.P., Cox, S.B., Cobb, G.P., and Anderson, T.A. (2006b). Evidence that miRNAs are different from other RNAs. Cell. Mol. Life Sci. **63,** 246-254.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31,** 3406-3415.

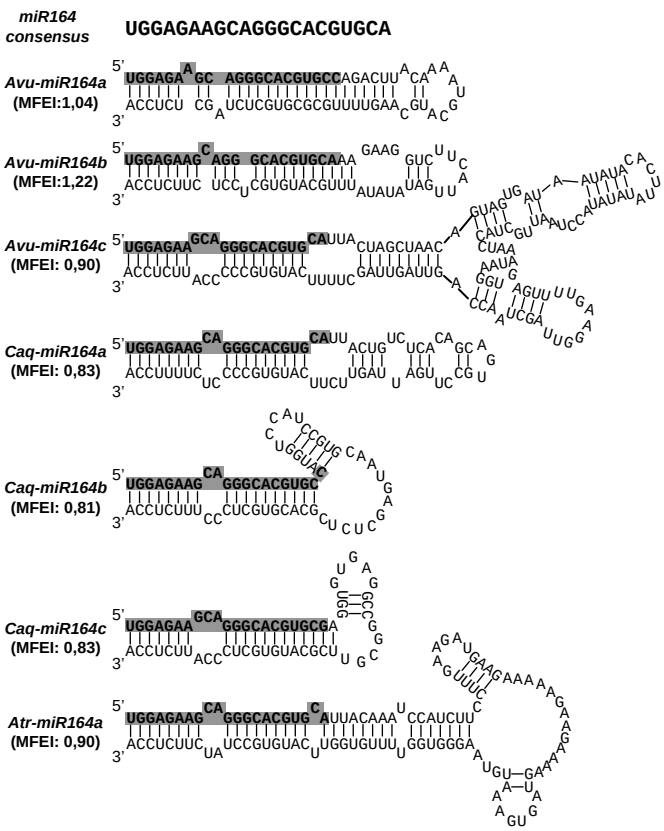


Figure 1. Secondary structures and Minimum Free Energy Indices (MFEI) of *pre-miR164s* predicted from *Aquilegia vulgaris* (*Avu-MIR164a-c*), *Cabomba aquatica* (*Caq-MIR164a-c*) and *Amborella trichopoda (Atr-MIR164a).*

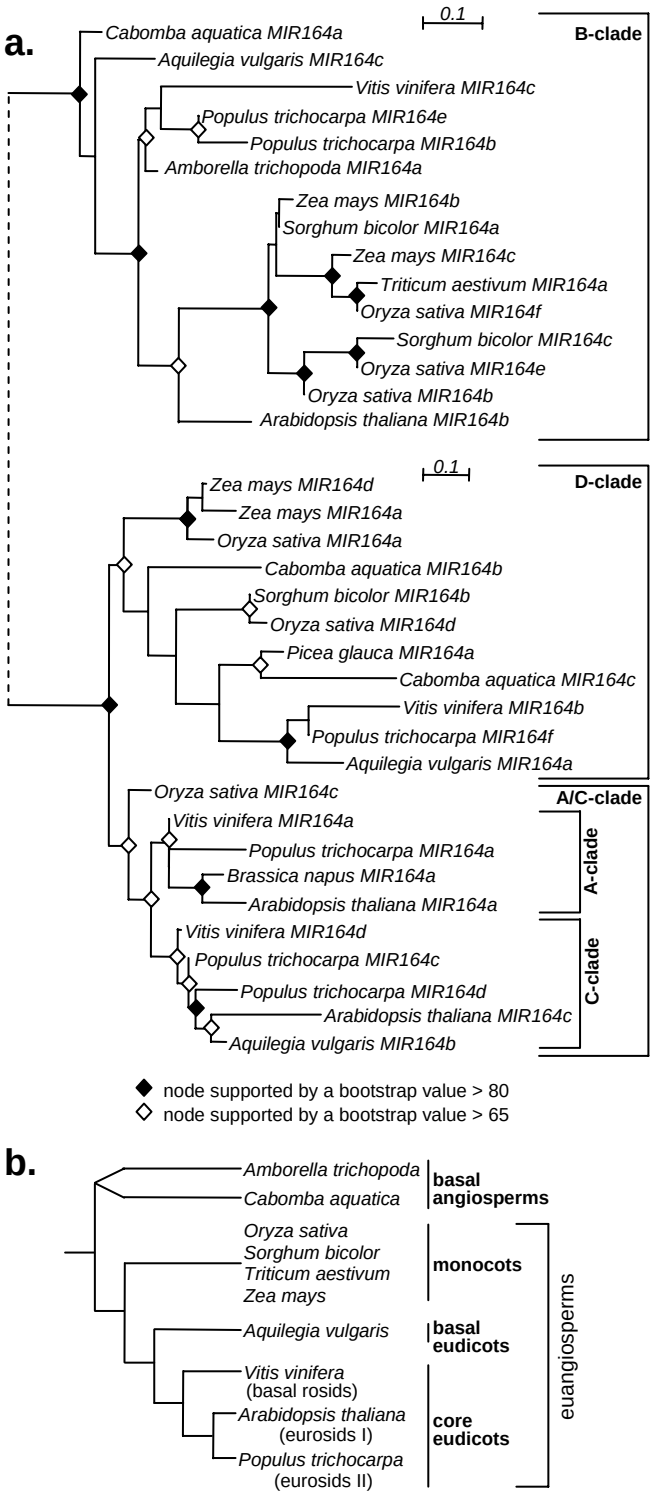


Figure 2. A phylogeny of the *MIR164* family in the angiosperms

a) A composite Maximum Likelihood phylogeny of the *MIR164* family, generated using Treefinder V.-May 2006 in two stages (two related phylogenies, linked by a dashed line). Nodes corresponding to high and moderate levels of bootstrap support are indicated. Accession numbers for the genes included are given in Suppl. Tab. 2.

b) The consensus view of angiosperm phylogeny (Stevens, 2001), showing the relationships of species sampled in the present study.

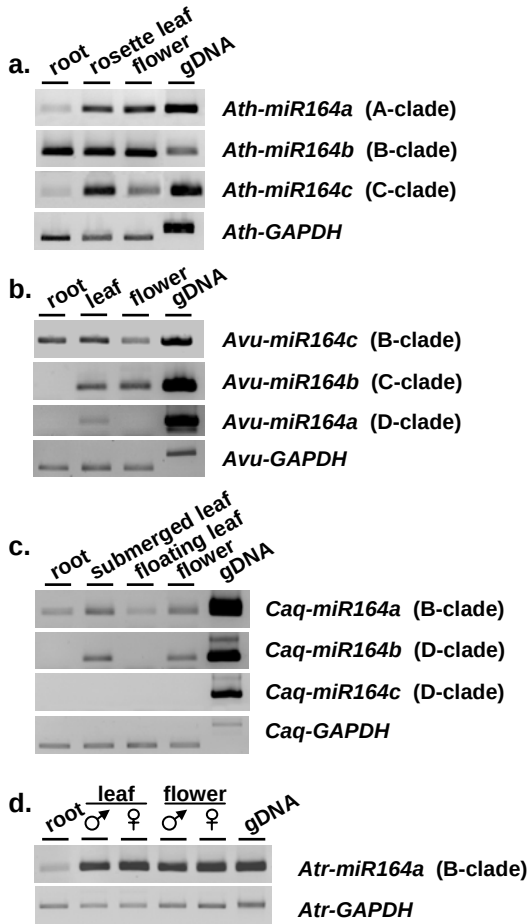


Figure 3. Semi-quantitative RT-PCR analysis of *MIR164* gene expression.
Amplifications of *miR164* and *GAPDH* transcripts were simultaneously performed on first-strand cDNA prepared from the tissues indicated of: (a) *Arabidopsis thaliana*, (b) *Aquilegia vulgaris,* (c) *Cabomba aquatica* and (d) *Amborella trichopoda*. Positive control amplifications from genomic DNA are also shown. Negative control amplifications, from RNA samples untreated with Reverse Transcriptase (not shown), failed in all cases to yield PCR products.

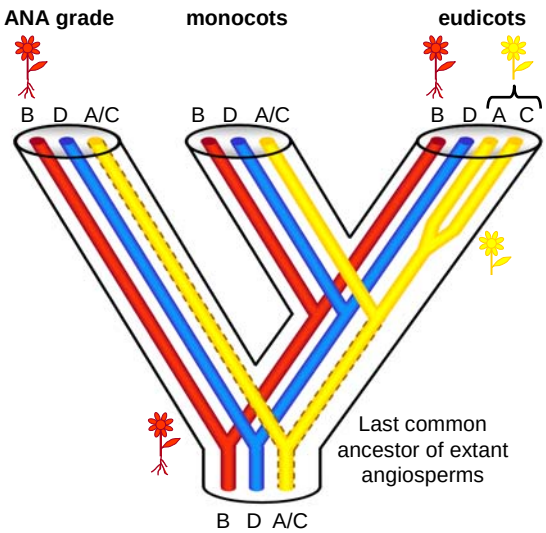
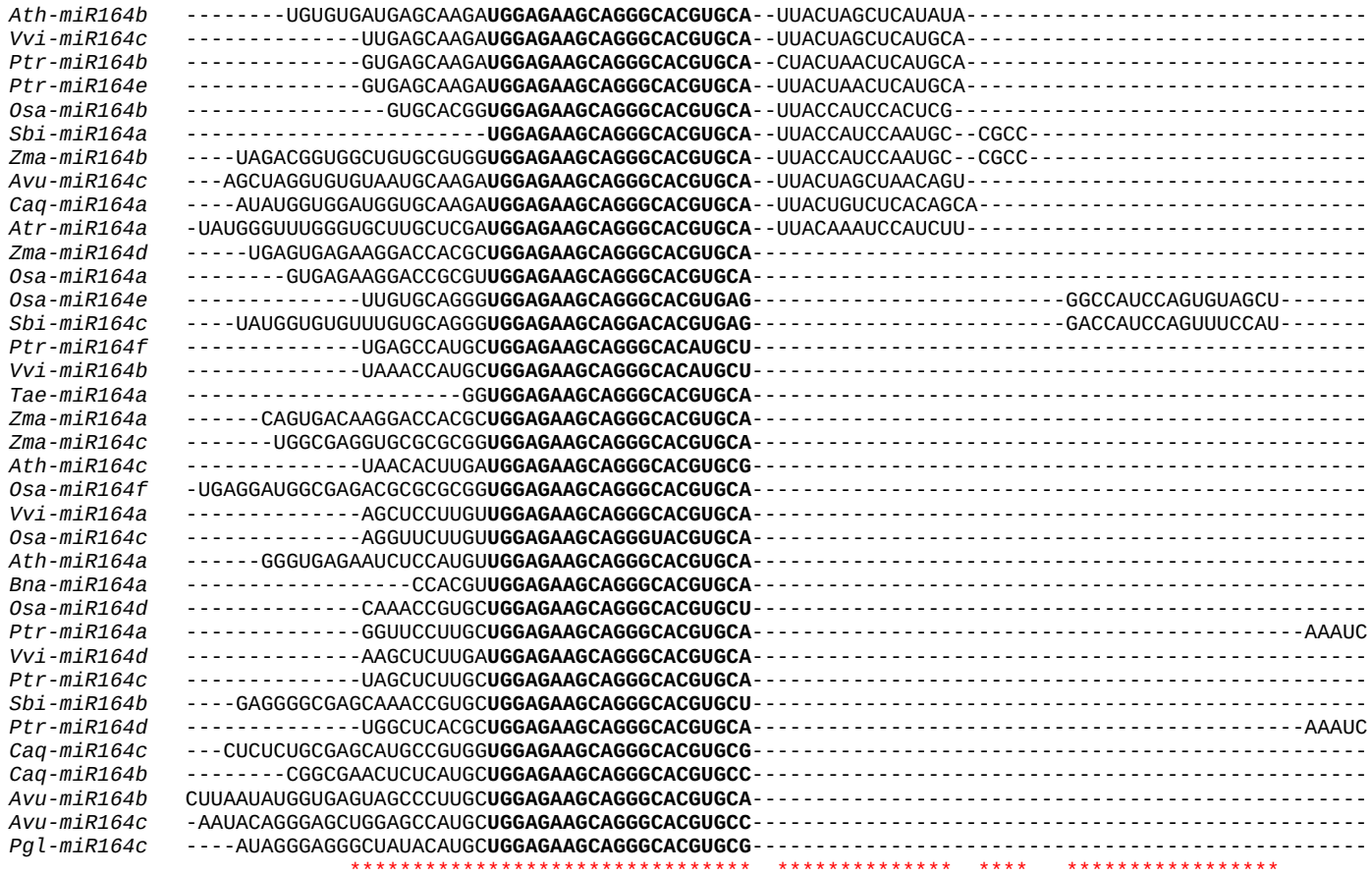
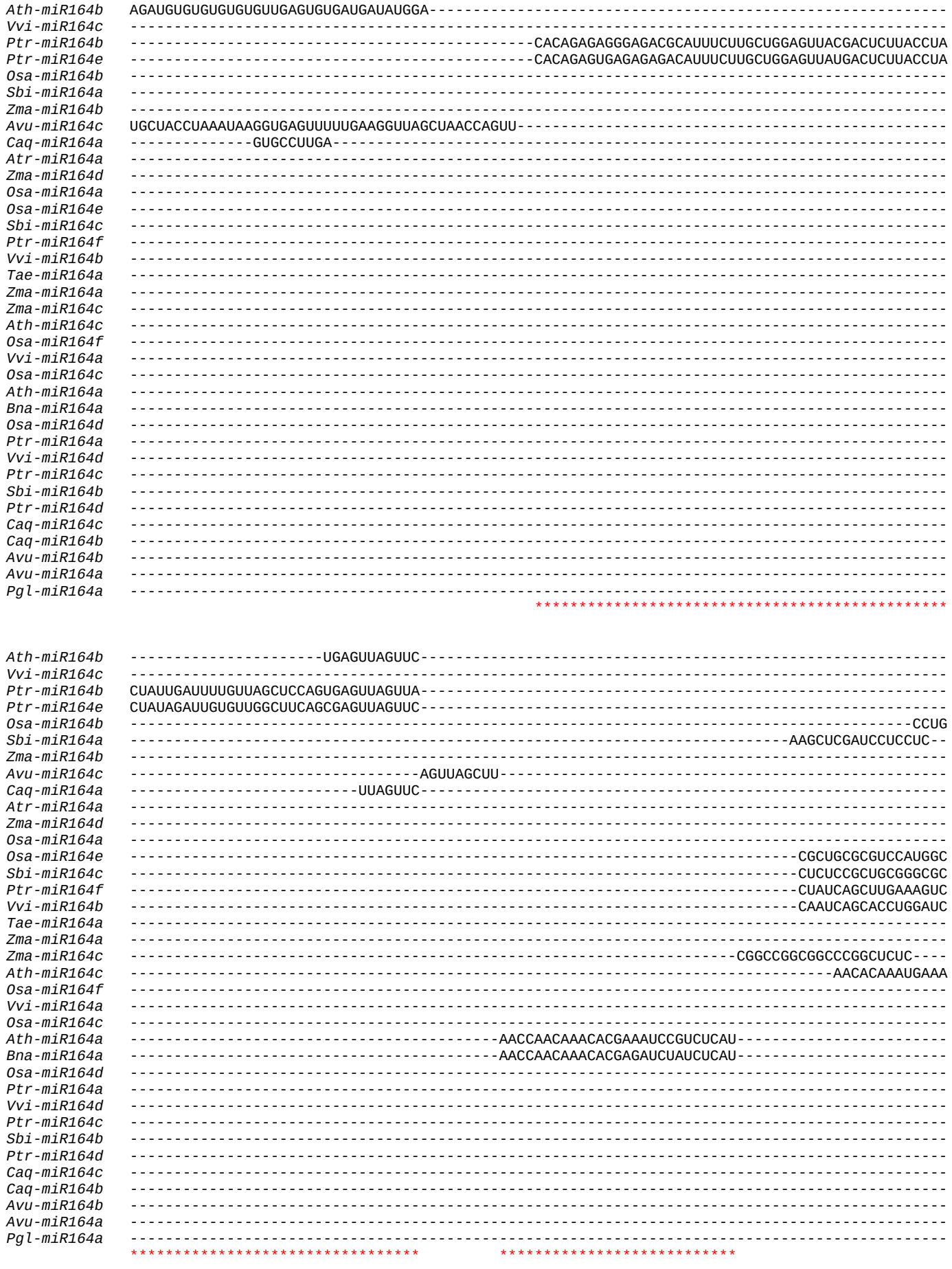


Figure 4. A partial reconstruction of *MIR164* family evolution in the angiosperms. Major lineages of *MIR164* genes are represented by colored bars, bordered by broken lines where evidence for their presence is indirect. Generalized expression patterns of *MIR164* genes in flowers, leaves and roots are indicated by plant symbols, matching the colors of the corresponding gene clades.

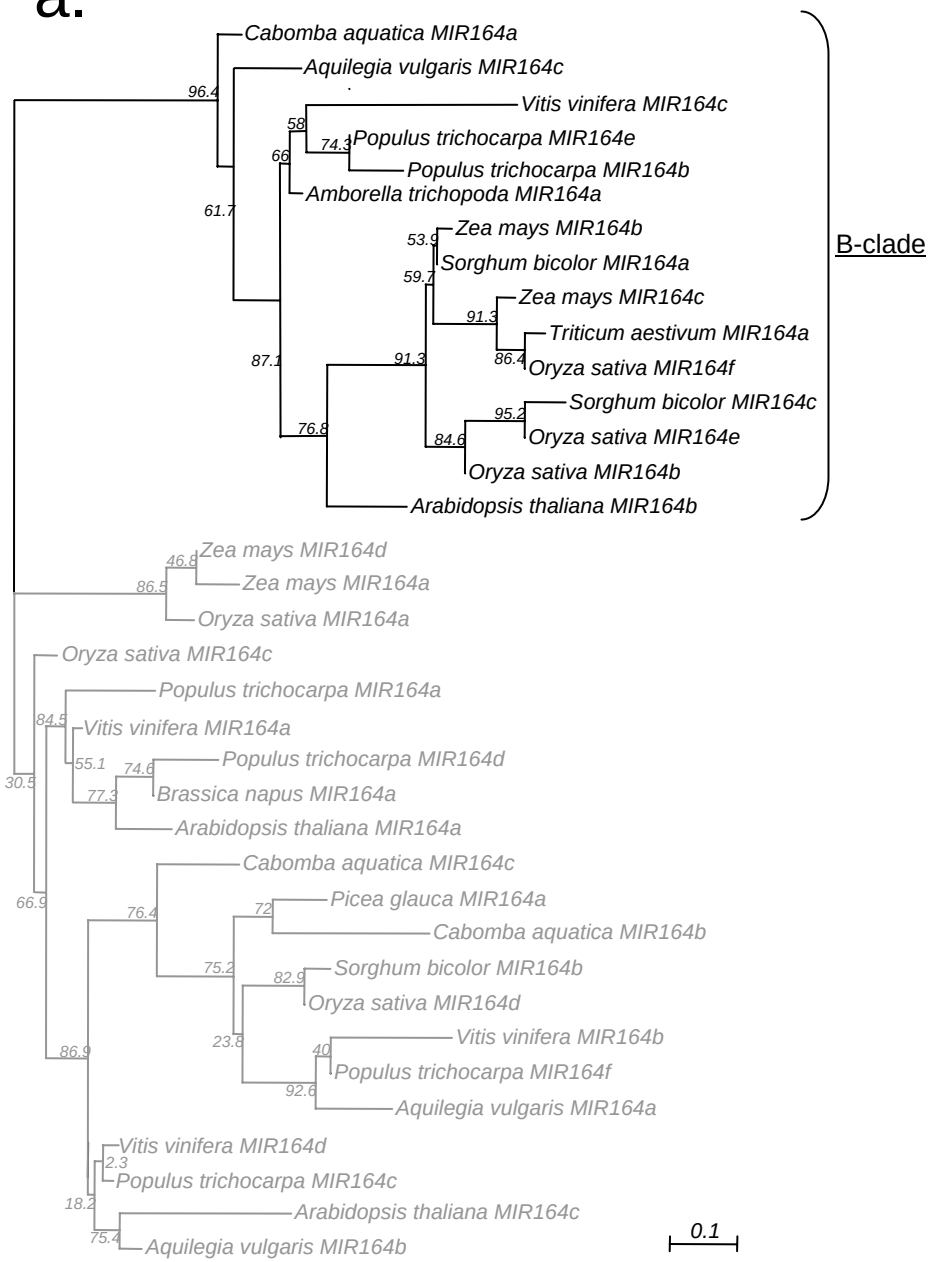

Ath-miR164b
Vvi-miR164c
Ptr-miR164b
Ptr-miR164e
Osa-miR164b
Sbi-miR164a
Zma-miR164b
Avu-miR164c
Caq-miR164a
Atr-miR164a
Zma-miR164d
Osa-miR164a
Osa-miR164e
Sbi-miR164c
Ptr-miR164f
Vvi-miR164b
Tae-miR164a
Zma-miR164a
Zma-miR164c
Ath-miR164c
Osa-miR164f
Vvi-miR164a
Osa-miR164c
Ath-miR164a
Bna-miR164a
Osa-miR164d
Ptr-miR164a
Vvi-miR164d
Ptr-miR164c
Sbi-miR164b
Ptr-miR164d
Caq-miR164c
Caq-miR164b
Avu-miR164b
Avu-miR164c
Pgl-miR164c

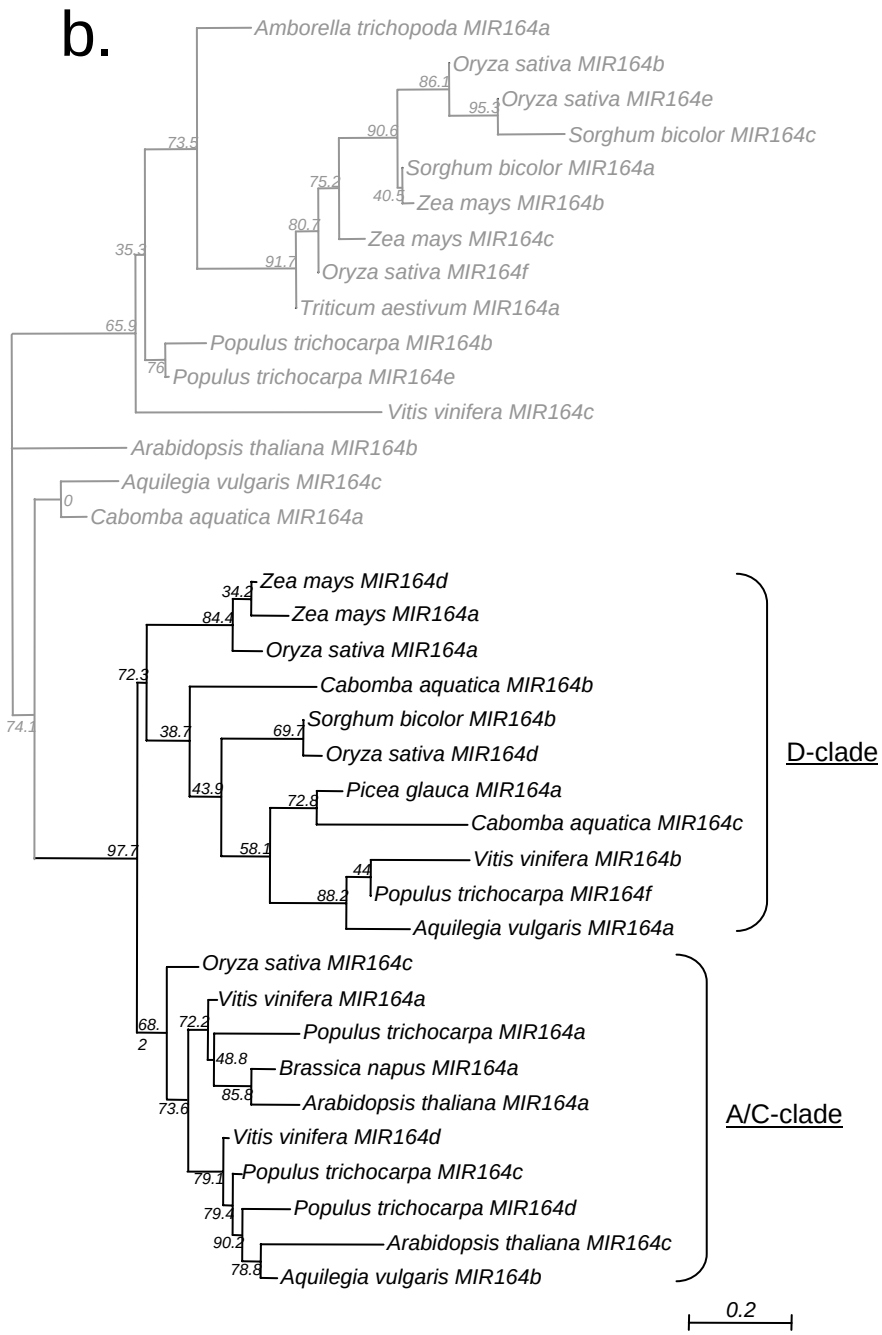

Ath-miR164b AGAUGUGUGUGUGUGUUGAGUGUGAUGAUAUGGA
Vvi-miR164c
Ptr-miR164b CACAGAGAGGGAGACGCAUUUCUUGCUGGAGUUACGACUCUUACCUA
Ptr-miR164e CACAGAGUGAGAGAGACAUUUCUUGCUGGAGUUAUGACUCUUACCUA
Osa-miR164b
Sbi-miR164a
Zma-miR164b
Avu-miR164c UGCUACCUAAAUAAGGUGAGUUUUUGAAGGUUAGCUAACCAGUU
Caq-miR164a GUGCCUUGA
Atr-miR164a
Zma-miR164d
Osa-miR164a
Osa-miR164e
Sbi-miR164c
Ptr-miR164f
Vvi-miR164b
Tae-miR164a
Zma-miR164a
Zma-miR164c
Ath-miR164c
Osa-miR164f
Vvi-miR164a
Osa-miR164c
Ath-miR164a
Bna-miR164a
Osa-miR164d
Ptr-miR164a
Vvi-miR164d
Ptr-miR164c
Sbi-miR164b
Ptr-miR164d
Caq-miR164c
Caq-miR164b
Avu-miR164b
Avu-miR164a
Pgl-miR164a
Ath-miR164b UGAGUUAGUUC
Vvi-miR164c
Ptr-miR164b CUAUUGAUUUUGUUAGCUCCAGUGAGUUAGUUA
Ptr-miR164e CUAUAGAUUGUGUUGGCUUCAGCGAGUUAGUUC
Osa-miR164b CCUG
Sbi-miR164a AAGCUCGAUCCUCCUC
Zma-miR164b
Avu-miR164c AGUUAGCUU
Caq-miR164a UUAGUUC
Atr-miR164a
Zma-miR164d
Osa-miR164a
Osa-miR164e CGCUGCGCGUCCAUGGC
Sbi-miR164c CUCUCCGCUGCGGGCGC
Ptr-miR164f CUAUCAGCUUGAAAGUC
Vvi-miR164b CAAUCAGCACCUGGAUC
Tae-miR164a
Zma-miR164a
Zma-miR164c CGGCCGGCGGCCCGGCUCUC
Ath-miR164c AACACAAAUGAAA
Osa-miR164f
Vvi-miR164a
Osa-miR164c
Ath-miR164a AACCAACAAACACGAAAUCCGUCUCAU
Bna-miR164a AACCAACAAACACGAGAUCUAUCUCAU
Osa-miR164d
Ptr-miR164a
Vvi-miR164d
Ptr-miR164c
Sbi-miR164b
Ptr-miR164d
Caq-miR164c
Caq-miR164b
Avu-miR164b
Avu-miR164a
Pgl-miR164a


Ath-miR164b
Vvi-miR164c
Ptr-miR164b
Ptr-miR164e
Osa-miR164b
Sbi-miR164a
Zma-miR164b
Avu-miR164c
Caq-miR164a
Atr-miR164a
Zma-miR164d
Osa-miR164a
Osa-miR164e
Sbi-miR164c
Ptr-miR164f
Vvi-miR164b
Tae-miR164a
Zma-miR164a
Zma-miR164c
Ath-miR164c
Osa-miR164f
Vvi-miR164a
Osa-miR164c
Ath-miR164a
Bna-miR164a
Osa-miR164d
Ptr-miR164a
Vvi-miR164d
Ptr-miR164c
Sbi-miR164b
Ptr-miR164d
Caq-miR164c
Caq-miR164b
Avu-miR164b
Avu-miR164a
Pgl-miR164a

Supplementary Figure 1. The multiple alignment of *MIR164* genes used as input in molecular phylogenetic analyses in the present work.

Sites used for phylogenetic analyses are indicated with asterisks. miR and miR* sequences are indicated in bold type.


a.
Cabomba aquatica MIR164a
Aquilegia vulgaris MIR164c
96.4
Vitis vinifera MIR164c
58
74.3
Populus trichocarpa MIR164e
66
Populus trichocarpa MIR164b
Amborella trichopoda MIR164a
61.7
Zea mays MIR164b
53.9
Sorghum bicolor MIR164a
59.7
Zea mays MIR164c
91.3
Triticum aestivum MIR164a
87.1
91.3
86.4
Oryza sativa MIR164f
Sorghum bicolor MIR164c
95.2
76.8
Oryza sativa MIR164e
84.6
Oryza sativa MIR164b
Arabidopsis thaliana MIR164b
B-clade
Zea mays MIR164d
46.8
86.5
Zea mays MIR164a
Oryza sativa MIR164a
Oryza sativa MIR164c
Populus trichocarpa MIR164a
84.5
Vitis vinifera MIR164a
55.1
74.6
Populus trichocarpa MIR164d
30.5
77.3
Brassica napus MIR164a
Arabidopsis thaliana MIR164a
Cabomba aquatica MIR164c
66.9
Picea glauca MIR164a
76.4
72
Cabomba aquatica MIR164b
75.2
82.9
Sorghum bicolor MIR164b
Oryza sativa MIR164d
23.8
Vitis vinifera MIR164b
86.9
40
Populus trichocarpa MIR164f
92.6
Aquilegia vulgaris MIR164a
Vitis vinifera MIR164d
2.3
Populus trichocarpa MIR164c
18.2
Arabidopsis thaliana MIR164c
75.4
Aquilegia vulgaris MIR164b
0.1



Supplementary Figure 2. The two *MIR164* phylogenies used to assemble the composite phylogeny in Fig 2a. The high and moderate bootstrap values give are summarized in Fig. 2a to indicate well- and moderately-supported nodes.

A first Maximum Likelihood phylogeny of *MIR164* genes, using Treefinder V.-May 2006 and assuming a GTR+Γ4 model. This analysis resolved the phylogeny of the B-clade (shown in black, see Fig 2a), but failed to adequately resolve the phylogeny of the A/C and D-clade sequences (shown in grey).

A second Maximum Likelihood phylogeny of *MIR164* genes from the dataset in SI Fig 5, using Treefinder V.-May 2006 and assuming a GTR+I+Γ4 model. This analysis resolved the phylogeny of the A/C- and D-clades (shown in black, see Fig 2a), but failed to adequately resolve the phylogeny of the B-clade sequences (shown in grey).

Plant species	Gene name	Number of recombinant clones analyzed	Number of positives	Frequency
Arabidopsis thaliana	*Ath-MIR164a*	520	40	8 %
	Ath-MIR164b		30	6 %
	Ath-MIR164c		12	2 %
Zea mays	*Zma-MIR164a*	1040	1	0.1 %
	Zma-MIR164b		8	0.8 %
	Zma-MIR164d		19	2 %
Aquilegia vulgaris	*Avu-MIR164a*	218	2	0.9 %
	Avu-MIR164b		1	0.5 %
	Avu-MIR164c		5	2 %
Cabomba aquatica	*Caq-MIR164a*	228	2	0.9 %
	Caq-MIR164b		1	0.4 %
	Caq-MIR164c		1	0.4 %
Amborella trichopoda	*Atr-MIR164a*	227	4	2 %

Supplementary Information-Table 1. Frequencies of *MIR164* genes identified.

Analyses were performed by colony hybridizations using gene-specific radioprobes for known *MIR164* genes from *Arabidopsis thaliana and Zea mays*, and by BLAST searching for miR* sites in novel *MIR164* genes from *Aquilegia vulgaris, Cabomba aquatica* and *Amborella trichopoda.*

Gene name	*MIR164* gene clade	Accession numbers in miRBase, EMBL*, and Genbank#
Amborella trichopoda MIR164a	B	FM162590*
Aquilegia vulgaris MIR164a	D	FM162584*
Aquilegia vulgaris MIR164b	C	FM162585*
Aquilegia vulgaris MIR164c	B	FM162586*
Arabidopsis thaliana MIR164a	A	MI0000197
Arabidopsis thaliana MIR164b	B	MI0000198
Arabidopsis thaliana MIR164c	C	MI0001087
Brassica napus MIR164a	A	MI0006478
Cabomba aquatica MIR164a	B	FM162587*
Cabomba aquatica MIR164b	D	FM162588*
Cabomba aquatica MIR164c	D	FM162589*
Oryza sativa MIR164a	D	MI0000668
Oryza sativa MIR164b	B	MI0000669
Oryza sativa MIR164c	A/C	MI0001103
Oryza sativa MIR164d	D	MI0001104
Oryza sativa MIR164e	B	MI0001105
Oryza sativa MIR164f	B	MI0001159
Picea glauca MIR164a	D	DR579554#
Populus trichocarpa MIR164a	A	MI0002212
Populus trichocarpa MIR164b	B	MI0002213
Populus trichocarpa MIR164c	C	MI0002214
Populus trichocarpa MIR164d	C	MI0002215
Populus trichocarpa MIR164e	B	MI0002216
Populus trichocarpa MIR164f	D	MI0002217
Sorghum bicolor MIR164a	B	MI0001512
Sorghum bicolor MIR164b	D	MI0001549
Sorghum bicolor MIR164c	B	MI0001852
Triticum aestivum MIR164a	B	MI0006173
Vitis vinifera MIR164a	A	MI0006503
Vitis vinifera MIR164b	D	MI0006504
Vitis vinifera MIR164c	B	MI0006505
Vitis vinifera MIR164d	C	MI0006506
Zea mays MIR164a	D	MI0001469
Zea mays MIR164b	B	MI0001471
Zea mays MIR164c	B	MI0001472
Zea mays MIR164d	D	MI0001470

Supplementary Information-Table 2. Accession numbers of *MIR164* genes used in phylogenetic analyses

Genes amplified	Forward and reverse primer sequences	Annealing temperature	Number of thermal cycles
Ath-MIR164a	5'-ACATGAGCTCTTCACCCATTG 5'-CCTTTTAGTTTCATGTGCATTG	55°C	45
Ath-MIR164b	5'-ATGACCACTCCACCTTGGTG 5'-CTGATACTATTTCGCAAGAC	55°C	45
Ath-MIR164c	5'-ACACGTGTCTCTCCCCCTCC 5'-ACTCGTATGCATTATTATTAC	55°C	45
Avu-MIR164a	5'-TCAATCAACTCTCTTCTTGGC 5'-CTGATGAGCAGTCAAATACTAGGG	60°C	55
Avu-MIR164b	5'-CCCACACCTTCTCTCCCTC 5'-AAACTGCTGCCTGACATACC	55°C	40
Avu-MIR164c	5'-CTAGTTCAATTCTTCCTTCC 5'-ATGGTTTGTACGTACCCAAGTG	55°C	40
Caq-MIR164a	5'-AATCTAAGGCATCACCCTTCC 5'-AAGAACTAATCAAGGCACTG	55°C	40
Caq-MIR164b	5'-CTGCCGCATTTCCCCTTC 5'-CGCGCATGATTTCCATTGAGAG	60°C	48
Caq-MIR164c	5'-GGAAGGAGGATCTGACGGTG 5'-GCTGATATTTGCCCTCATCCAGAC	60°C	48
Atr-MIR164a	5'-GCTTGGGGAGGCAACGATGC 5'-GGTAGAGCTTGTAATCTCCACTGATC	58°C	45
Ath-GAPDH	5'-AGGGTGGTGCCAAGAAGGTTG 5'-GTAGCCCCACTCGTTGTCGTA	55°C	25
Caq-GAPDH		55°C	27
Avu-GAPDH		53°C	25
Atr-GAPDH		58°C	33

Supplementary Information-Table 3. Primer sequences and annealing temperatures used in Reverse Transcriptase PCR analyses.

III) Perspectives

Cette étude présente une méthodologie pour cloner les gènes MIR et reconstruire la phylogénie de la famille génique chez n'importe quelle espèce dont le génome n'est pas séquencé. Cependant, rappelons que ce travail se base sur l'exemple de la seule famille des gènes *MIR164* chez les plantes à fleurs. Or, on peut se demander si (i) la méthode s'applique à d'autres familles de gènes MIR, (ii) la divergence (en terme de séquence nucléotidique) ne sera pas trop grande en dehors des angiospermes, et notamment entre les plantes à fleurs et les gymnospermes. De surcroît, la méthode de clonage s'est révélée non exhaustive chez les espèces du grade ANA, rendant ainsi délicat l'inférence du nombre de gènes chez l'ancêtre commun des plantes à fleurs. Certes, cette difficulté peut être contournée en augmentant le nombre d'espèces échantillonnées, mais ceci ne va pas sans un lourd travail expérimental. Par conséquent, il serait intéressant d'automatiser ce protocole afin d'en faire une méthode moyen- ou haut-débit.

La méthode développée dans le manuscrit ci-dessus utilise l'exemple spécifique de la famille des gènes *MIR164* chez les plantes à fleurs. Il serait intéressant de voir si cette méthode est applicable aux animaux et champignons. En effet, la conservation de signatures moléculaires découvertes entre les domaines miR et miR* des gènes *MIR164* dénote une contrainte fonctionnelle. Cette contrainte fonctionnelle peut agir au niveau de la biosynthèse des microARNs ou bien du mode d'action de ces petits ARNs sur leurs gènes cibles. Or, nous savons par exemple que les microARNs des animaux régulent l'expression de leurs gènes cibles en bloquant la traduction alors que chez les plantes, la régulation se fait essentiellement par clivage de l'ARN messager du gène cible.

CONCLUSIONS - PERSPECTIVES

Chapitre 6 : CONCLUSIONS – PERSPECTIVES

Finalement, où en sommes-nous avec l'apparition des plantes à fleurs ? Faut-il être aussi radical et négatif que le titre de la dernière revue de Michael Frohlich et Mark Chase (Frohlich and Chase, 2007): "After a dozen years of progress the origin of angiosperms is still a great mystery" ?

La théorie "Mostly Male" est de plus en plus remise en question. Une étude récente de l'expression des gènes *LFY* et *NEEDLY* chez trois genres de conifères *lato sensu* montre qu'ils sont exprimés à la fois dans les cônes mâles et femelles (Vazquez-Lobo et al., 2007). Ces données semblent aller contre l'idée de l'expression sexe spécifique de chacun de ces deux paralogues. En parallèle, la théorie "Mostly Male" postule que le rapprochement des deux sexes résulte de l'apparition d'ovules ectopiques sur la microsporophylle. Vérifier cette hypothèse revient donc à montrer que les gènes du développement du carpelle sont majoritairement exprimés dans les cônes mâles des gymnospermes actuelles. Cette hypothèse n'est pas soutenue par une analyse statistique des données d'expression des différents orthologues chez les gymnospermes des gènes impliqués dans le développement du carpelle des angiospermes (Raquel Tavares, communication personnelle). N'oublions pas que cette théorie ne s'est jamais voulue comme définitive mais plutôt comme hypothèse de travail.

Un des points que je voudrais également souligner est le manque de données génomiques concernant les gymnospermes. Actuellement, quatre génomes nucléaires complets sont disponibles chez les plantes à fleurs : *A. thaliana*, *O. sativa*, *P. trichocarpa* et *V. vinifera*. Dans les prochaines années, nous allons assister à une explosion de génomes nucléaires chez les angiospermes avec pas moins de 24 projets de séquençage en cours (voir liste complète à : http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Search&db=genomeprj&term="Embryophyta"%5BOrganism%5D). Concernant les gymnospermes, la liste de projets se résume pour le moment à la seule espèce *Pinus taeda*. Notons que cette espèce représentera un travail conséquent de séquençage et surtout d'assemblage avec une taille de génome de 1C = 21658 Mpb (pour comparaison avec *Zea mays* dont l'assemblage n'est pas trivial : 1C = 2671 Mpb). L'obtention du génome nucléaire d'une gymnosperme constituera un véritable atout pour les approches de type "évo-dévo" dont le raisonnement repose beaucoup sur le contenu en gène des différents génomes. En effet, c'est une manière de formuler des hypothèses évolutives quant à l'acquisition de nouvelles fonctions au cours de l'évolution. En d'autres termes, c'est

une manière de combler le manque d'outils fonctionnels disponibles chez les espèces non-modèles en génétique du développement.

Un moyen de tester la conservation ou non de la fonction d'un gène et d'inférer la fonction ancestrale de ce lignage serait de développer de nouveaux modèles génétiques. Ces nouveaux modèles devront répondre à différents critères tels qu'une position phylogénétique d'intérêt au sein des spermaphytes (pour répondre à une certaine problématique initiale), aisément cultivables au laboratoire, et pour lesquels la transformation génétique soit efficace. Actuellement, une technique de transformation comme le VIGS (**V**irus-**I**nduced **G**ene **S**ilencing) semble constituer une bonne alternative à la transformation stable. Cette technique consiste à infecter la plante avec des virus porteurs de séquences du gène à réprimer (Baulcombe, 1999). Cette méthode a déjà apporté des résultats chez des espèces de plantes à fleurs dont la position phylogénétique est intéressante d'un point de vue évolutif : *Aquilegia vulgaris* (Gould and Kramer, 2007), *Eschscholzia californica* (Hileman et al., 2005). Cependant, ces espèces n'ont pas une position phylogénétique aussi basale que les espèces du grade ANA. C'est la raison pour laquelle notre équipe de recherche envisage à long terme de développer *Cabomba aquatica* (Nymphéales) comme espèce modèle.

***Cabomba aquatica* comme espèce modèle.** La taille de son génome a été estimée à 2C = 6,74 pg (soit $6{,}6.10^9$ pb) alors qu'une étude statistique réalisée sur 79 angiospermes basales révèle une taille génomique moyenne de 2C = 4,42 ± 4,4 pg (Bennett and Leitch, 2004). Ainsi, le génome nucléaire de *C. aquatica* ne présente pas une taille singulière au sein des Angiospermes basales. De même, les Angiospermes basales ne présentent pas des tailles de génomes significativement différentes de celles des Eudicots avec en moyenne 2C = 6,28 ± 6,28 pg (Bennett and Leitch, 2004). Néanmoins, la taille du génome de *C. aquatica* est utile sur un plan technique. En effet, cette valeur de 2C = 6,74 pg rend possible de futures expériences de southern blot mais dissuade de réaliser une banque génomique. Concernant les études d'expression, l'hybridation *in situ* fonctionne déjà chez cette espèce. En conclusion, *C. aquatica* serait sans doute le meilleur système modèle à développer dans le grade ANA. Ses inconvénients sont en effet mineurs si on considère que la plupart des espèces du grade ANA sont ligneuses (*Amborella, Iliicium, Drimys*), compliquant fortement la transformation génétique de ces dernières. Le manque d'outils génétiques, incontournables pour l'approche fonctionnelle, est sans doute le facteur le plus limitant pour des approches comparatives complètes.

Cependant, le plus important reste la prise de conscience que les espèces d'études en biologie (*Arabidopsis, Caenorhabditis, Danio, Drosophila, Mus, Saccharomyces*, etc…) ne constituent que des modèles dont le développement est singulier et partiellement dérivé. De plus en plus de travaux soulignent des différences majeures entre le développement des espèces modèles et des modèles émergents : *Drosophila vs. Nasonia* (Brent et al., 2007), *Arabidopsis vs. Zea* (Chandler et al., 2008; Nardmann and Werr, 2007). Les différents projets de ma thèse soulignent l'importance d'un échantillonnage large et représentatif de la diversité des organismes. Ceci constitue le meilleur moyen de distinguer les caractères ancestraux des caractères dérivés, que ce soit au niveau morphologique ou moléculaire. Ainsi, le clonage des orthologues des gènes *ARF3* et *ARF4* chez plusieurs espèces du grade ANA a permis de montrer que la perte des domaines III et IV pouvait résulter de mécanismes différents. L'étude phylogénomique de la famille ARF a permis de relativiser l'importance macro-évolutive de la régulation via les uORFs. Cette régulation semble spécifique à certaines espèces d'angiospermes récentes (core eudicots). De manière plus rare, des études menées chez des espèces non modèles ont permis de découvrir des choses fondamentales qui n'avaient pas été décelées chez des espèces modèles, et qui ont été généralisées à ces dernières dans un second temps. Juste pour donner un exemple récent, il a été montré que les courts introns du génome de *Paramecium tetraurelia* dont la taille est multiple de 3 (dans le cadre du lecture de la protéine codée) étaient sélectionnés de telle sorte à posséder des codons-stop afin de prévenir des éventuels rétentions d'introns lors de l'épissage (Jaillon et al., 2008). Ce résultat a été par la suite généralisé à d'autres espèces dont le génome est séquencé comme *A. thaliana*, *Homo sapiens*, *D. melanogaster* (Jaillon et al., 2008).

BIBLIOGRAPHIE

Albert, V. A., Oppenheimer, D. G. and Lindqvist, C. (2002). Pleiotropy, redundancy and the evolution of flowers. *Trends Plant Sci* **7**, 297-301.
Alvarez, J. and Smyth, D. R. (1999). CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development* **126**, 2377-86.
Axtell, M. J. and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell* **17**, 1658-73.
Baker, C. C., Sieber, P., Wellmer, F. and Meyerowitz, E. M. (2005). The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. *Curr Biol* **15**, 303-15.
Balanza, V., Navarrete, M., Trigueros, M. and Ferrandiz, C. (2006). Patterning the female side of Arabidopsis: the importance of hormones. *J Exp Bot* **57**, 3457-69.
Barakat, A., Wall, K., Leebens-Mack, J., Wang, Y. J., Carlson, J. E. and Depamphilis, C. W. (2007). Large-scale identification of microRNAs from a basal eudicot (Eschscholzia californica) and conservation in flowering plants. *Plant J* **51**, 991-1003.
Barkman, T. J., Chenery, G., McNeal, J. R., Lyons-Weiler, J., Ellisens, W. J., Moore, G., Wolfe, A. D. and dePamphilis, C. W. (2000). Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proc Natl Acad Sci U S A* **97**, 13166-71.
Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281-97.
Baulcombe, D. C. (1999). Fast forward genetics based on virus-induced gene silencing. *Curr Opin Plant Biol* **2**, 109-13.
Baum, D. A. and Hileman, L. C. (2006). A genetic model for the origin of flowers. *Flowering and its manipulation.*
Becker, A., Kaufmann, K., Freialdenhoven, A., Vincent, C., Li, M. A., Saedler, H. and Theissen, G. (2002). A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. *Mol Genet Genomics* **266**, 942-50.
Becker, A., Winter, K. U., Meyer, B., Saedler, H. and Theissen, G. (2000). MADS-Box gene diversity in seed plants 300 million years ago. *Mol Biol Evol* **17**, 1425-34.
Benedito, V. A., Visser, P. B., van Tuyl, J. M., Angenent, G. C., de Vries, S. C. and Krens, F. A. (2004). Ectopic expression of LLAG1, an AGAMOUS homologue from lily (Lilium longiflorum Thunb.) causes floral homeotic modifications in Arabidopsis. *J Exp Bot* **55**, 1391-9.
Bennett, M. and Leitch, I. (2004). Angiosperm DNA C-values database (release 5.0, Dec. 2004).
Bennett, S. R. M., Alvarez, J., Bossinger, G. and Smyth, D. R. (1995). Morphogenesis in pinoid mutants of Arabidopsis thaliana. *The Plant Journal* **8**, 505-520.
Blazquez, M. A. and Weigel, D. (2000). Integration of floral inductive signals in Arabidopsis. *Nature* **404**, 889-92.
Bowman, J. L. (2000). Axial patterning in leaves and other lateral organs. *Curr Opin Genet Dev* **10**, 399-404.
Bowman, J. L., Alvarez, J., Weigel, D., Meyerowitz, E. M. and Smyth, D. R. (1993). Control of flower development in Arabidopsis thaliana by APETALA1 and interacting genes. *Development* **119**, 721-743.
Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1989). Genes directing flower development in Arabidopsis. *Plant Cell* **1**, 37-52.
Bowman, J. L., Smyth, D. R. and Meyerowitz, E. M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. *Development* **112**, 1-20.

Brent, A. E., Yucel, G., Small, S. and Desplan, C. (2007). Permissive and instructive anterior patterning rely on mRNA localization in the wasp embryo. *Science* **315**, 1841-3.
Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L. and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185-90.
Busch, M. A., Bomblies, K. and Weigel, D. (1999). Activation of a floral homeotic gene in Arabidopsis. *Science* **285**, 585-7.
Byzova, M. V., Franken, J., Aarts, M. G., de Almeida-Engler, J., Engler, G., Mariani, C., Van Lookeren Campagne, M. M. and Angenent, G. C. (1999). Arabidopsis STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. *Genes Dev* **13**, 1002-14.
Chandler, J., Nardmann, J. and Werr, W. (2008). Plant development revolves around axes. *Trends Plant Sci* **13**, 78-84.
Chaw, S. M., Parkinson, C. L., Cheng, Y., Vincent, T. M. and Palmer, J. D. (2000). Seed plant phylogeny inferred from all three plant genomes: monophyly of extant gymnosperms and origin of Gnetales from conifers. *Proc Natl Acad Sci U S A* **97**, 4086-91.
Chen, X., Liu, J., Cheng, Y. and Jia, D. (2002). HEN1 functions pleiotropically in Arabidopsis development and acts in C function in the flower. *Development* **129**, 1085-94.
Cheng, Y., Dai, X. and Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. *Genes Dev* **20**, 1790-9.
Cheng, Y., Kato, N., Wang, W., Li, J. and Chen, X. (2003). Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in Arabidopsis thaliana. *Dev Cell* **4**, 53-66.
Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J., Angenent, G. C. and van Tunen, A. J. (1995). The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell* **7**, 1859-68.
Crane, P. R. (1985) Phylogenetics analysis of seed plants and the origin of angiosperms. *Ann Missouri Bot Gard* **72**, 716-793.
Delsuc, F., Brinkmann, H. and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* **6**, 361-75.
Ditta, G., Pinyopich, A., Robles, P., Pelaz, S. and Yanofsky, M. F. (2004). The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. *Curr Biol* **14**, 1935-40.
Donoghue, M. J. and Doyle, J. A. (2000). Seed plant phylogeny: Demise of the anthophyte hypothesis? *Curr Biol* **10**, R106-9.
Dornelas, M. C., Van Lammeren, A. A. and Kreis, M. (2000). Arabidopsis thaliana SHAGGY-related protein kinases (AtSK11 and 12) function in perianth and gynoecium development. *Plant J* **21**, 419-29.
Doyle, J. A. and Donoghue, M. J. (1986). Seed plant phylogeny and the origin of angiosperms : an experimental cladistic approach. Bot Rev **52**, 321-431.
Drews, G. N., Bowman, J. L. and Meyerowitz, E. M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell* **65**, 991-1002.
Endress, P. K. and Igersheim, A. (2000). Gynoecium structure and evolution in basal angiosperms. *International Journal of Plant Science* **161**, S211-S223.
Fourquin, C., Vinauger-Douard, M., Chambrier, P., Berne-Dedieu, A. and Scutt, C. P. (2007). Functional conservation between CRABS CLAW orthologues from widely diverged angiosperms. *Ann Bot (Lond)* **100**, 651-7.
Fourquin, C., Vinauger-Douard, M., Fogliani, B., Dumas, C. and Scutt, C. P. (2005). Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel

development since the ancestor of the extant angiosperms. *Proc Natl Acad Sci U S A* **102**, 4649-54.
Friedman, W. E. (1990). Double Fertilization in Ephedra, a Nonflowering Seed Plant: Its Bearing on the Origin of Angiosperms. *Science* **247**, 951-954.
Friedman, W. E. and Floyd, S. K. (2001). Perspective: the origin of flowering plants and their reproductive biology--a tale of two phylogenies. *Evolution Int J Org Evolution* **55**, 217-31.
Friis, E. M., Doyle, J. A., Endress, P. K. and Leng, Q. (2003). Archaefructus--angiosperm precursor or specialized early angiosperm? *Trends Plant Sci* **8**, 369-73.
Friis, E. M., Pedersen, K. R. and Crane, P. R. (1999). Early angiosperm diversification: the diversity of pollen associated with angiosperm reproductive structures in Early Cretaceous floras from Portugal. *Annals Missouri Bot. Gard.* **86**, 259-296.
Friis, E. M., Pedersen, K. R. and Crane, P. R. (2000). Reproductive structure and organization of basal angiosperms from the Early Cretaceius (Barremian or Aptian) of western Portugal. *International Journal of Plant Science* **161**, S169-S182.
Friis, E. M., Pedersen, K. R. and Crane, P. R. (2001). Fossil evidence of water lilies (Nymphaeales) in the Early Cretaceous. *Nature* **410**, 357-60.
Friis, E. M., Pedersen, K. R. and Crane, P. R. (2005). When Earth started blooming: insights from the fossil record. *Curr Opin Plant Biol* **8**, 5-12.
Frohlich, M. W. (2003). An evolutionary scenario for the origin of flowers. *Nat Rev Genet* **4**, 559-66.
Frohlich, M. W. and Chase, M. W. (2007). After a dozen years of progress the origin of angiosperms is still a great mystery. *Nature* **450**, 1184-9.
Frohlich, M. W. and Meyerowitz, E. M. (1997). The search for flower homeotic gene homologs in basal angiosperms and Gnetales: a potential new source of data on the evolutionary origin of flowers. *International Journal of Plant Science* **158**, S131-S142.
Frohlich, M. W. and Parker, D. S. (2000). The mostly male theory of flower evolutionary origins: from genes to fossils. *Systematic Botany* **25**, 155-170.
Goremykin, V. V., Hirsch-Ernst, K. I., Wolfl, S. and Hellwig, F. H. (2003). Analysis of the Amborella trichopoda chloroplast genome sequence suggests that amborella is not a basal angiosperm. *Mol Biol Evol* **20**, 1499-505.
Goremykin, V. V., Hirsch-Ernst, K. I., Wolfl, S. and Hellwig, F. H. (2004). The chloroplast genome of Nymphaea alba: whole-genome analyses and the problem of identifying the most basal angiosperm. *Mol Biol Evol* **21**, 1445-54.
Gould, B. and Kramer, E. M. (2007). Virus-induced gene silencing as a tool for functional analyses in the emerging model plant Aquilegia (columbine, Ranunculaceae). *Plant Methods* **3**, 6.
Graham, L. E. (1993). Origin of land plants. *(Wiley, New York)*.
Graham, S. W. and Olmstead, R. G. (2000). Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *Am J Bot* **87**, 1712-1730.
Graybeal, A. (1998). Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst Biol* **47**, 9-17.
Groth-Malonek, M., Pruchner, D., Grewe, F. and Knoop, V. (2005). Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol Biol Evol* **22**, 117-25.
Gustafson-Brown, C., Savidge, B. and Yanofsky, M. F. (1994). Regulation of the arabidopsis floral homeotic gene APETALA1. *Cell* **76**, 131-43.
Hileman, L. C., Drea, S., Martino, G., Litt, A. and Irish, V. F. (2005). Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species Papaver somniferum (opium poppy). *Plant J* **44**, 334-41.

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**, 525-9.
Jaillon, O., Bouhouche, K., Gout, J. F., Aury, J. M., Noel, B., Saudemont, B., Nowacki, M., Serrano, V., Porcel, B. M., Segurens, B. et al. (2008). Translational control of intron splicing in eukaryotes. *Nature* **451**, 359-62.
Jang, S., Hur, J., Kim, S. J., Han, M. J., Kim, S. R. and An, G. (2004). Ectopic expression of OsYAB1 causes extra stamens and carpels in rice. *Plant Mol Biol* **56**, 133-43.
Juarez, M. T., Twigg, R. W. and Timmermans, M. C. (2004). Specification of adaxial cell fate during maize leaf development. *Development* **131**, 4533-44.
Kang. (1995). Phenotypic alterations of petal and sepal by ectopic expression of a rice MADS box gene in tobacco. *Plant Mol Biol* **29**, 1-10.
Kempin, S. A., Savidge, B. and Yanofsky, M. F. (1995). Molecular basis of the cauliflower phenotype in Arabidopsis. *Science* **267**, 522-5.
Klavins, S. D., Taylor, T. N. and Taylor, E. L. (2002). Anatomy of Umkomasia (Corystospermales) from the triassic of Antarctica. *American Journal of Botany* **89**, 664-676.
Korall, P., Pryer, K. M., Metzgar, J. S., Schneider, H. and Conant, D. S. (2006). Tree ferns: monophyletic groups and their relationships as revealed by four protein-coding plastid loci. *Mol Phylogenet Evol* **39**, 830-45.
Kramer, E. M. and Irish, V. F. (2000). Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. *International Journal of Plant Science* **161**, S29-S40.
Kramer, E. M., Jaramillo, M. A. and Di Stilio, V. S. (2004). Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166**, 1011-23.
Krizek, B. A., Prost, V. and Macias, A. (2000). AINTEGUMENTA promotes petal identity and acts as a negative regulator of AGAMOUS. *Plant Cell* **12**, 1357-66.
Lee, J. Y., Baum, S. F., Oh, S. H., Jiang, C. Z., Chen, J. C. and Bowman, J. L. (2005). Recruitment of CRABS CLAW to promote nectary development within the eudicot clade. *Development* **132**, 5021-32.
Leebens-Mack, J., Raubeson, L. A., Cui, L., Kuehl, J. V., Fourcade, M. H., Chumley, T. W., Boore, J. L., Jansen, R. K. and depamphilis, C. W. (2005). Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol Biol Evol* **22**, 1948-63.
Lenhard, M., Bohnert, A., Jurgens, G. and Laux, T. (2001). Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between WUSCHEL and AGAMOUS. *Cell* **105**, 805-14.
Lewis, L. A., Mishler, B. D. and Vilgalys, R. (1997). Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene rbcL. *Mol Phylogenet Evol* **7**, 377-93.
Li, J. and Chen, X. (2003). PAUSED, a putative exportin-t, acts pleiotropically in Arabidopsis development but is dispensable for viability. *Plant Physiol* **132**, 1913-24.
Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S. and Yanofsky, M. F. (1999). Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *Plant Cell* **11**, 1007-18.
Liu, Z. and Meyerowitz, E. M. (1995). LEUNIG regulates AGAMOUS expression in Arabidopsis flowers. *Development* **121**, 975-91.
Llave, C., Xie, Z., Kasschau, K. D. and Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**, 2053-6.

Lohmann, J. U., Hong, R. L., Hobe, M., Busch, M. A., Parcy, F., Simon, R. and Weigel, D. (2001). A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* **105**, 793-803.
Mallory, A. C. and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. *Nat Genet* **38 Suppl**, S31-6.
Mandel, M. A., Bowman, J. L., Kempin, S. A., Ma, H., Meyerowitz, E. M. and Yanofsky, M. F. (1992). Manipulation of flower structure in transgenic tobacco. *Cell* **71**, 133-43.
Mellerowicz, E. J., Horgan, K., Walden, A., Coker, A. and Walter, C. (1998). PRFLL--a Pinus radiata homologue of FLORICAULA and LEAFY is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. *Planta* **206**, 619-29.
Mishler, B. D. and Churchill, S. P. (1984). A cladistic approach to the phylogeny of the "Bryophytes". *Brittonia* **36**, 406-424.
Mishler, B. D. and Churchill, S. P. (1985). Transition to a land flora: phylogenetic relationships of the green algae and bryophytes. *Cladistics* **1**, 305-328.
Mishler, B. D., Lewis, L. A., Buchheim, M. A., Renzaglia, K. S., Garbary, D. J., Delwiche, C. F., Zechman, F. W., Kantz, T. S. and Chapman, R. L. (1994). Phylogenetic relationships of the "green algae" and "bryophytes". *Ann. Missouri Bot. Gard.* **81**, 451-483.
Mizukami, Y. and Ma, H. (1992). Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. *Cell* **71**, 119-31.
Moore, M. J., Bell, C. D., Soltis, P. S. and Soltis, D. E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci U S A* **104**, 19363-8.
Mouradov, A., Glassick, T., Hamdorf, B., Murphy, L., Fowler, B., Marla, S. and Teasdale, R. D. (1998). NEEDLY, a Pinus radiata ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. *Proc Natl Acad Sci U S A* **95**, 6537-42.
Nardmann, J. and Werr, W. (2007). The evolution of plant regulatory networks: what Arabidopsis cannot say for itself. *Curr Opin Plant Biol* **10**, 653-9.
Nemhauser, J. L., Feldman, L. J. and Zambryski, P. C. (2000). Auxin and ETTIN in Arabidopsis gynoecium morphogenesis. *Development* **127**, 3877-88.
Nickrent, D. L., Parkinson, C. L., Palmer, J. D. and Duff, R. J. (2000). Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol Biol Evol* **17**, 1885-95.
Nishiyama, T., Wolf, P. G., Kugita, M., Sinclair, R. B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K. et al. (2004). Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol Biol Evol* **21**, 1813-9.
Okada, K., Ueda, J., Komaki, M. K., Bell, C. J. and Shimura, Y. (1991). Requirement of the Auxin Polar Transport System in Early Stages of Arabidopsis Floral Bud Formation. *Plant Cell* **3**, 677-684.
Parcy, F., Bomblies, K. and Weigel, D. (2002). Interaction of LEAFY, AGAMOUS and TERMINAL FLOWER1 in maintaining floral meristem identity in Arabidopsis. *Development* **129**, 2519-27.
Peaucelle, A., Morin, H., Traas, J. and Laufs, P. (2007). Plants expressing a miR164-resistant CUC2 gene reveal the importance of post-meristematic maintenance of phyllotaxy in Arabidopsis. *Development* **134**, 1045-50.
Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E. and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405**, 200-3.
Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E. and Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85-8.

Qiu, Y. L., Cho, Y., Cox, J. C. and Palmer, J. D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394**, 671-4.
Qiu, Y. L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Zimmer, E. A., Chen, Z., Savolainen, V. and Chase, M. W. (1999). The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402**, 404-7.
Qiu, Y. L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J. et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci U S A* **103**, 15511-6.
Ratcliffe, O. J., Bradley, D. J. and Coen, E. S. (1999). Separation of shoot and floral identity in Arabidopsis. *Development* **126**, 1109-20.
Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P. F., Lindquist, E. A., Kamisugi, Y. et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64-9.
Rutherford, G., Tanurdzic, M., Hasebe, M. and Banks, J. A. (2004). A systemic gene silencing method suitable for high throughput, reverse genetic analyses of gene function in fern gametophytes. *BMC Plant Biol* **4**, 6.
Saarela, J. M., Rai, H. S., Doyle, J. A., Endress, P. K., Mathews, S., Marchant, A. D., Briggs, B. G. and Graham, S. W. (2007). Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* **446**, 312-5.
Saporta, G. D. (1894). Flore fossile du Portugal. Nouvelles contributions à la flore Mésozoique. Accompagnées d'une notice stratigraphique par Paul Choffat. *Imprimerie de l'Académie Royale des Sciences, Lisbon.*
Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M. and Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. *Dev Cell* **8**, 517-27.
Sieber, P., Wellmer, F., Gheyselinck, J., Riechmann, J. L. and Meyerowitz, E. M. (2007). Redundancy and specialization among plant microRNAs: role of the MIR164 family in developmental robustness. *Development* **134**, 1051-60.
Sohlberg, J. J., Myrenas, M., Kuusk, S., Lagercrantz, U., Kowalczyk, M., Sandberg, G. and Sundberg, E. (2006). STY1 regulates auxin homeostasis and affects apical-basal patterning of the Arabidopsis gynoecium. *Plant J* **47**, 112-23.
Soltis, D. E. and Soltis, P. S. (2004). Amborella not a "basal angiosperm" ? Not so fast. *American Journal of Botany* **91**, 997-1001.
Stebbins, G. L. (1974). Flowering plants: evolution above the species level. *(Harvard Univ. Press, Cambridge, Massachusetts).*
Stefanovic, S., Rice, D. W. and Palmer, J. D. (2004). Long branch attraction, taxon sampling, and the earliest angiosperms: Amborella or monocots? *BMC Evol Biol* **4**, 35.
Sun, G., Dilcher, D. L., Zheng, S. and Zhou, Z. (1998). In search of the first flower: A jurassic angiosperm, archaefructus, from northeast china. *Science* **282**, 1692-5.
Sun, G., Ji, Q., Dilcher, D. L., Zheng, S., Nixon, K. C. and Wang, X. (2002). Archaefructaceae, a new basal angiosperm family. *Science* **296**, 899-904.
Surange, K. R. and Chandra, S. (1971). Denkania indica gen. et sp. nov. - a glossopteridean fructification from the lower Gondwana of India. *Palaeobotanist* **20**, 264-268.
Tang, G., Reinhart, B. J., Bartel, D. P. and Zamore, P. D. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev* **17**, 49-63.
Theissen, G. and Becker, A. (2004). Gymnosperm orthologues of class B floral homeotic genes and their impact on understanding flower origin. *Critical Reviews in Plant Sciences.* **23**, 129-148.
Theissen, G., Becker, A., Di Rosa, A., Kanno, A., Kim, J. T., Munster, T., Winter, K. U. and Saedler, H. (2000). A short history of MADS-box genes in plants. *Plant Mol Biol* **42**, 115-49.

Toriba, T., Harada, K., Takamura, A., Nakamura, H., Ichikawa, H., Suzaki, T. and Hirano, H. Y. (2007). Molecular characterization the YABBY gene family in Oryza sativa and expression analysis of OsYABBY1. *Mol Genet Genomics* **277**, 457-68.
Turmel, M., Otis, C. and Lemieux, C. (2006). The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. *Mol Biol Evol* **23**, 1324-38.
Ulmasov, T., Hagen, G. and Guilfoyle, T. J. (1999). Dimerization and DNA binding of auxin response factors. *Plant J* **19**, 309-19.
Vandenbussche, M., Theissen, G., Van de Peer, Y. and Gerats, T. (2003). Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Res* **31**, 4401-9.
Vazquez-Lobo, A., Carlsbecker, A., Vergara-Silva, F., Alvarez-Buylla, E. R., Pinero, D. and Engstrom, P. (2007). Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera Picea, Podocarpus, and Taxus: implications for current evo-devo hypotheses for gymnosperms. *Evol Dev* **9**, 446-59.
Wagner, D., Sablowski, R. W. and Meyerowitz, E. M. (1999). Transcriptional activation of APETALA1 by LEAFY. *Science* **285**, 582-4.
Wang, W. and Chen, X. (2004). HUA ENHANCER3 reveals a role for a cyclin-dependent protein kinase in the specification of floral organ identity in Arabidopsis. *Development* **131**, 3147-56.
Wang, X., Duan, S., Geng, B., Cui, J. and Yang, Y. (2007). Schmeissneria: a missing link to angiosperms? *BMC Evol Biol* **7**, 14.
Winter, K. U., Becker, A., Munster, T., Kim, J. T., Saedler, H. and Theissen, G. (1999). MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc Natl Acad Sci U S A* **96**, 7342-7.
Wolf, P. G., Karol, K. G., Mandoli, D. F., Kuehl, J., Arumuganathan, K., Ellis, M. W., Mishler, B. D., Kelch, D. G., Olmstead, R. G. and Boore, J. L. (2005). The first complete chloroplast genome sequence of a lycophyte, Huperzia lucidula (Lycopodiaceae). *Gene* **350**, 117-28.
Yamada, T., Ito, M. and Kato, M. (2003). Expression pattern of INNER NO OUTER homologue in Nymphaea (water lily family, Nymphaeaceae). *Dev Genes Evol* **213**, 510-3.
Yamaguchi, T., Nagasawa, N., Kawasaki, S., Matsuoka, M., Nagato, Y. and Hirano, H. Y. (2004). The YABBY gene DROOPING LEAF regulates carpel specification and midrib development in Oryza sativa. *Plant Cell* **16**, 500-9.
Yao, X. L., Taylor, T. N. and Taylor, E. L. (1995). The corystosperm pollen organ Pteruchus from the triassic of antarctica. *American Journal of Botany* **82**, 535-546.
Yu, H., Ito, T., Wellmer, F. and Meyerowitz, E. M. (2004). Repression of AGAMOUS-LIKE 24 is a crucial step in promoting flower development. *Nat Genet* **36**, 157-61.
Zanis, M. J., Soltis, D. E., Soltis, P. S., Mathews, S. and Donoghue, M. J. (2002). The root of the angiosperms revisited. *Proc Natl Acad Sci U S A* **99**, 6848-53.

ANNEXES

Journal of Experimental Botany, Vol. 57, No. 10, pp. 2143–2152, 2006
doi:10.1093/jxb/erj188 Advance Access publication 23 May, 2006

Journal of Experimental Botany

REVIEW ARTICLE

An evolutionary perspective on the regulation of carpel development

Charlie P. Scutt*, Marion Vinauger-Douard, Chloé Fourquin, Cédric Finet and Christian Dumas

Laboratoire de Reproduction et Développement des Plantes, Unité Mixte de Recherche 5667- CNRS - INRA - ENS- Université Lyon 1 - and Institut Fédératif de Recherche 128, Ecole Normale Supérieure de Lyon, 46, allée d'Italie, F-69364 Lyon Cedex 07, France

Received 15 December 2005; Accepted 9 March 2006

Abstract

The carpel, or female reproductive organ enclosing the ovules, is one of the major evolutionary innovations of the flowering plants. The control of carpel development has been intensively studied in the model eudicot species *Arabidopsis thaliana*. This review traces the evolutionary history of genes involved in carpel development by surveying orthologous genes in taxa whose lineages separated from that of *A. thaliana* at different levels of the phylogenetic tree of the seed plants. Some aspects of the control of female reproductive development are conserved between the flowering plants and their sister group, the gymnosperms, indicating the presence of these in the common ancestor of the extant seeds plants, some 300 million years ago. Gene duplications that took place in the pre-angiosperm lineage, before the evolution of the first flowering plants, provided novel gene clades of potential importance for the origin of the carpel. Subsequent to the appearance of the first flowering plants, further gene duplications have led to sub-functionalization events, in which pre-existing reproductive functions were shared between paralogous gene clades. In some cases, fluidity in gene function is evident, leading to similar functions in carpel development being controlled by non-orthologous genes in different taxa. In other cases, gene duplication events have created sequences that evolved novel functions by the process of neo-functionalization, thereby generating biodiversity in carpel and fruit structures.

Key words: Angiosperms, carpel, development, evolution, flower, flowering-plants, gynoecium, pistil.

The big cover up

In the gymnosperms, the most ancient group of living seed plants, ovules most frequently occur as naked structures that develop in the axils of leaf-like organs. By contrast, in the more recently evolved flowering plants or angiosperms, the ovules are enclosed and protected by a specialized female reproductive organ termed the carpel. Besides protecting the ovules, the carpel confers numerous further advantages on the flowering plants. Stigma tissues at the carpel apex are adapted in different species for the efficient capture of pollen carried by vectors including insects, mammals, birds, and the wind. In addition, the carpel provides a location for selective mechanisms that operate on pollen, such as self-incompatibility, which promotes out-breeding. Following pollination, compatible pollen tubes are guided with meticulous accuracy through the tissues of the carpel, specifically toward unfertilized ovules. After fertilization, the carpel tissues undergo further developmental changes to become the fruit, which protects the developing seeds and later contributes to the dissemination of these by a wide variety of mechanisms in different species. For all of these reasons, the carpel was undoubtedly a major factor in the evolutionary success of the angiosperms, which diversified from a common ancestor that is estimated to have lived in the Late Jurassic period, around 160 million years ago (MYA), to form approximately 300 000 species alive today.

The molecular control of carpel development has been investigated in several model species, although most thoroughly in *Arabidopsis thaliana* of the Brassicaceae. In parallel, molecular phylogenetic studies have now clarified the evolutionary relationships between the major groups of seed plants (Fig. 1), as reviewed by Kuzoff and Gasser (2000). The combination of developmental and

* To whom correspondence should be addressed. E-mail: Charlie.Scutt@ens-lyon.fr

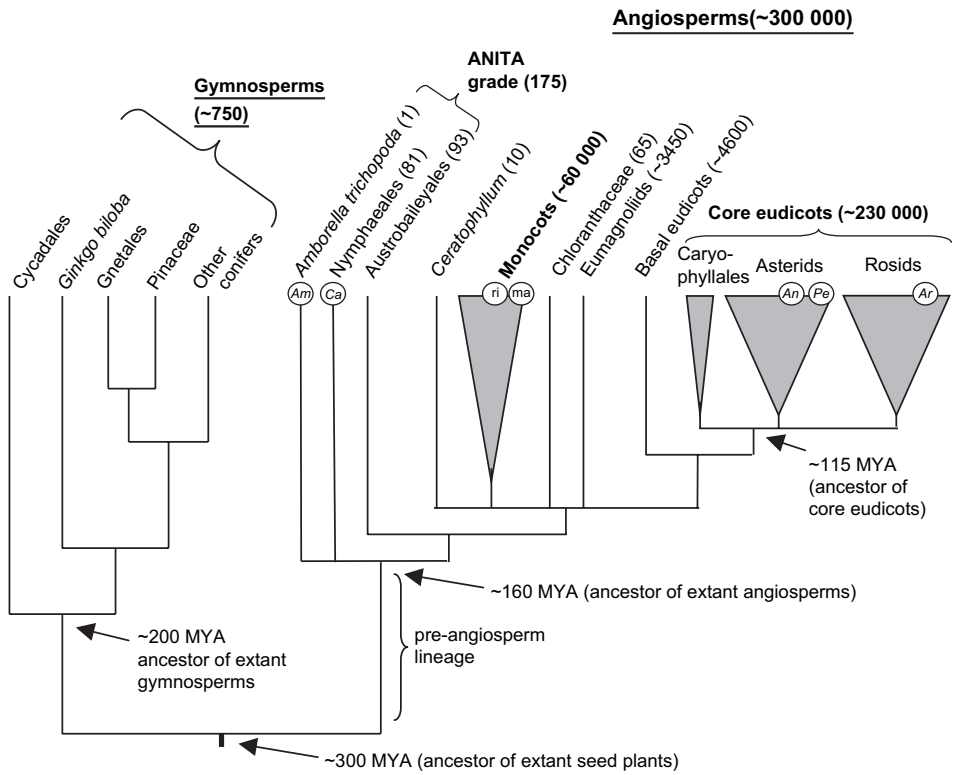


Fig. 1. The phylogeny of the seed plants, based on a consensus of molecular phylogenetic studies. The numbers of species in major clades are given in parentheses, while approximate dates of divergence are taken from Davies *et al.* (2004), based on a calibration of the molecular clock using fossil data. Very large clades are represented by shaded triangles. The positions of certain species referred to in the text are indicated as follows: *Am*, *Amborella trichopoda*; *An*, *Antirrhinum majus*; *Ar*, *Arabidopsis thaliana*; *Ca*, *Cabomba aquatica*; ma, maize; *Pe*, *Petunia hybrida*; ri, rice.

phylogenetic information provides a starting point to unravel the evolution of carpel development from the pre-angiosperm lineage through to present day model species such as *A. thaliana*. In addition, the comparison of carpel development mechanisms in different extant angiosperm groups should allow the identification of the molecular differences that underlie the diversity of carpel and fruit morphology throughout the flowering plants.

Before the carpel

The extant gymnosperms have been shown by molecular phylogenetic analyses to form a monophyletic group in a sister position to the angiosperms (Fig. 1). By the comparative analysis of reproductive development in gymnosperms and angiosperms, something may be deduced of the molecular mechanisms of female development that existed before the carpel. The ABC model for the development of a typical angiosperm flower (Coen and Meyerowitz, 1991), postulates a 'C-function' to specify carpel development in the fourth floral whorl (Fig. 2a). This model further postulates the combination of C-function activity with that of a 'B-function' to specify stamen development in the third whorl. The genes encoding the B- and C-functions have been identified from several model angiosperms and found to encode MADS box transcription factors of the Type II MIKC class (Parenicova *et al.*, 2003). Analyses of taxa from the major gymnosperm groups: Pinaceae (Tandre *et al.*, 1995), Gnetales (Becker *et al.*, 2000), Ginkgoales (Jager *et al.*, 2003), and Cycadales (Zhang *et al.*, 2004), clearly indicate the presence of both B- and C-function orthologues in gymnosperms. Male and female reproductive structures in gymnosperms develop on separate reproductive axes (cones etc), or even on separate individuals. C-function orthologues are expressed in both male and female reproductive axes in gymnosperms, whereas B-function orthologues are male-specific, mirroring the organ-specific expression of the equivalent B- and C-function genes in angiosperms (Fig. 2a, b). In addition, coding sequences of B- and C-function genes from gymnosperms show activities similar to those of the equivalent *A. thaliana* genes in transgenic *A. thaliana* (Tandre *et al.*, 1998; Winter *et al.*, 2002; Zhang *et al.*, 2004). It therefore appears that the last common ancestor of the extant seed plants, living some 300 MYA, possessed a C-function-like gene that played a role in the development of both male and female reproductive organs. The differentiation between the sexes in that ancestral seed plant would have depended on the male-specific expression of a B-function-like gene.

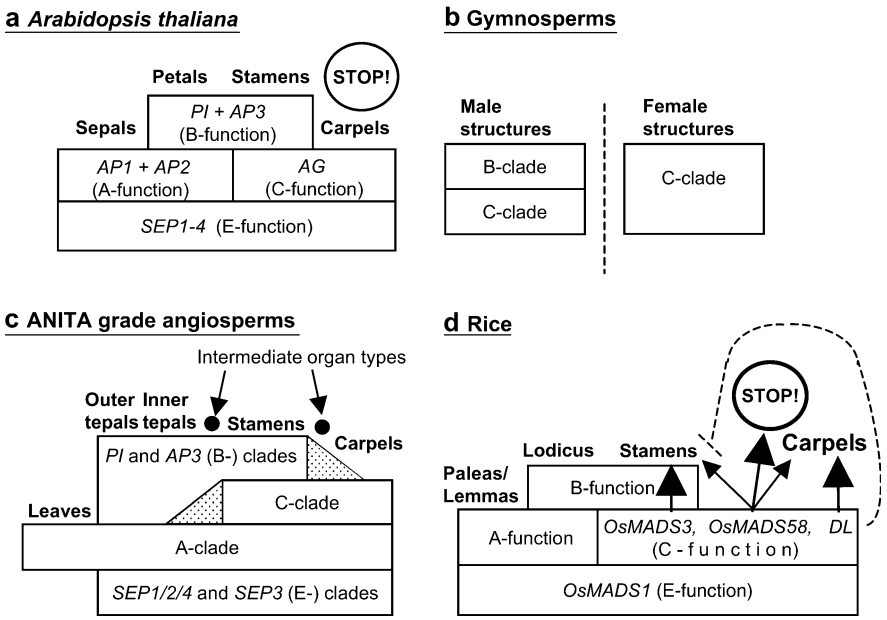


Fig. 2. The ABCE model of flower development in *A. thaliana*, and its derivatives in other taxa. (a) In *A. thaliana*, A-, B-, C-, and E-function floral homeotic genes, expressed in overlapping domains (horizontal bars) of the floral meristem, control the identities of floral organs in a combinatorial manner: A+E specifies sepal development in the first whorl, A+B+E specifies petal development in the second whorl, B+C+E specifies stamen development in the third whorl, and C+E specifies carpel development in the fourth whorl. In addition, the C-function causes an arrest of organ proliferation (the 'STOP' function) in the fourth whorl. (*AG*, *AGAMOUS*; *AP1*, *APETALA1*; *AP2*, *APETALA2*; *AP3*, *APETALA3*; *PI*, *PISTILLATA*; *SEP1–4*, *SEPALLATA1–4*.) (b) In gymnosperms, B- and C-clade MADS box genes are expressed in a combinatorial manner in male (B+C) and female (C alone) reproductive structures, resembling the expression of their *A. thaliana* orthologues in male and female floral organs. (c) In ANITA grade angiosperms, B- and C-clade MADS box gene expression resembles that of the respective *A. thaliana* orthologues, although with less well-defined boundaries (dotted areas). Strong B-clade gene expression is generally detected in the outer floral whorl of ANITA angiosperms, possibly reflecting an absence of developmental differentiation between whorls 1 and 2. A-clade MADS box gene expression differs radically between ANITA angiosperms and *A. thaliana*, extending throughout the flower and into leaves. (d) In rice flowers, typifying the Poaceae of the monocot clade, A-, B-, and E-function genes are expressed in similar patterns to those of their *A. thaliana* orthologues to specify specialized perianth organs (paleas, lemmas, and lodicules) and stamens. Two paralogous C-clade MADS box genes show a partial sub-functionalization between the third and fourth whorls, with one paralogue playing a major role in stamen development in the third whorl, while the other plays a major role in the 'stop' function in the fourth whorl (thick arrows, major roles; thin arrows, minor roles). The YABBY gene *DROOPING LEAF* (*DL*) plays a major role in carpel specification that is independent of C-clade MADS box gene expression. *DL* may act directly on carpel development (solid arrow), or indirectly by limiting the inner boundary of B-function gene expression (dashed arrow), or both of these.

In *A. thaliana*, B- and C-function genes have been shown to function together with a further class of MADS box genes encoding an 'E-function', thereby extending the ABC model to an ABCE model (Pelaz *et al.*, 2000; Honma and Goto, 2001). Accordingly, carpel development in *A. thaliana* requires a combination of activities of the C-function gene, *AGAMOUS* (*AG*), with that of the E-function, which is encoded by four genes, termed *SEPALLATA1–4* (*SEP1–4*), with extensively overlapping functions (Pelaz *et al.*, 2000; Ditta *et al.*, 2004). C- and E-function proteins are thought to act as hetero-tetramers (Theissen and Saedler, 2001) to control the transcription of their downstream target genes (Ito *et al.*, 2004; Gomez-Mena *et al.*, 2005) and thereby bring about carpel development.

E-clade genes have not been found in gymnosperms, but the closely related *AGAMOUS-LIKE6* (*AGL6*) MADS box clade is present in both angiosperms and gymnosperms (Carlsbecker *et al.*, 2004). These data suggest that a gene duplication event, generating the ancestors of the *AGL6* and *SEP* (E-clade) genes, occurred prior to the separation of the pre-angiosperm and gymnosperm lineages around 300 MYA (Becker and Theissen, 2003; Zahn *et al.*, 2005). If the E-clade did predate the ancestor of the extant seed plants as proposed, the flower (including the carpel) would not have evolved as a direct result of the origin of E-clade genes. However, the crucial mechanistic importance of the E-function for flower development in extant angiosperms implies that the recruitment, at least, of E-clade genes to these functions may have played a central role in the origin of this structure.

The extant angiosperms and gymnosperms both possess MADS box genes of a paralogous clade to the B-clade, termed B-sister genes (Becker *et al.*, 2002). Unlike the male-expressed B-function, B-sister genes seem to be expressed in female reproductive tissues and this characteristic is conserved between angiosperms and gymnosperms. The unique *A. thaliana* B-sister gene, *TRANSPARENT TESTA16*, plays a role in the pigmentation of the outer ovule integument (Nesi *et al.*, 2002), although it has been hypothesized that the widespread conservation of the

B-sister lineage is evidence of a more important ancestral role, probably in ovule development (Kaufmann *et al.*, 2005).

Theories for carpel origin

Carpels, along with the other principal floral organs, have for long been postulated to be modified from a leaf ground plan. Relatively recent experimental evidence supports this view: floral organs are converted to leaves in plants in which all of the A, B and C function genes (Coen and Meyerowitz, 1991), or the redundant E-function genes (Pelaz *et al.*, 2000), are inactivated. In addition, the ectopic expression of combinations of A, B or C with *SEP* (E-function) genes will convert leaves into floral organs (Honma and Goto, 2001).

Although the carpel appears to be a modified leaf, it may be more directly related to sporophylls, or leaves that carry spore-producing organs. As the carpel is female, it has traditionally been regarded as derived from megasporophylls that would have subtended ovules in the pre-angiosperm lineage. Accordingly, the carpel would be directly homologous to such gymnosperm organs as the female cone scales of conifers. A recent molecular explanation for the origin of the bisexual axis in the flowering plants, termed the Out-of-Male/Out-of-Female Theory (Theissen and Becker, 2004), is broadly consistent with this view of a female origin for the carpel. This theory proposes a pair of alternative mechanisms, based on the movement of a frontier of B-function gene expression in either a basipetal or acropetal direction along male or female reproductive axes, respectively, in the pre-angiosperm lineage. As a result, the axis affected is proposed to have become bisexual, with female organs at its tip and male organs at its base. Carpels would then have evolved by the closure of megasporophylls in the apical region of the bisexual axis.

Conversely, the 'Mostly Male Theory' (Frohlich and Parker, 2000; Frohlich, 2003) proposes the carpel to have been derived by the closure of (male) microsporophylls, around ovules that had developed ectopically on these. According to this view, all or most of the female-specific developmental pathways in the pre-angiosperm lineage, other than those required for ovule development, were lost during the evolution of the first angiosperms. One gene that was apparently lost prior to the radiation of the angiosperms is called *NEEDLY* (*NLY*). *NLY* is a gymnosperm-specific paralogue of *LEAFY* (*LFY*), which itself is present in all seed plants and is known to regulate positively B- and C-function genes in *A. thaliana*. Early studies suggested that *NLY* may specifically control female developmental programmes in gymnosperms, providing support for the Mostly Male Theory (Mouradov *et al.*, 1998). However, the sex-specific expression of *LFY* and *NLY* does not appear to be general in the gymnosperms (Carlsbecker *et al.*, 2004; Dornelas and Rodriguez, 2005). Although *LFY* and *NLY* may prove to be of lesser importance for the Mostly-Male Theory than was originally thought, it is possible that a systematic analysis of gene orthology and expression data between angiosperms and gymnosperms will provide other genes that could be used to test this and other theories that seek to explain the origin of the flower and carpel.

The ancestral carpel

Molecular phylogenetic analyses have clearly identified the first diverging lineages within the angiosperm clade (Mathews and Donoghue, 1999; Parkinson *et al.*, 1999; Qiu *et al.*, 1999; Soltis *et al.*, 1999; Barkman *et al.*, 2000). These are grouped into only three extant orders, Amborellales, Nymphaeales, and Austrobaileyales, collectively termed the ANITA grade. Amborellales contains the single species *Amborella trichopoda*, a small tree endemic to the tropical island of New Caledonia in the Southern Pacific. Nymphaeales is a cosmopolitan order containing two families of aquatic plants. Austrobaileyales contains four families, representing a mixture of endemic and more widely distributed groups. There is very good evidence that Amborellales and Nymphaeales diverged from the remaining angiosperm lineage before the divergence of Austrobaileyales (Aoki *et al.*, 2004; Stellari *et al.*, 2004). However, the relative order of divergence of the two most basal lineages, Amborellales and Nymphaeales, remains unclear. Most recent molecular phylogenies place Amborellales alone in the most basal position (Zanis *et al.*, 2002), while others group it together with Nymphaeales in a first-diverging clade (Qiu *et al.*, 2001).

Comparison of the features of ANITA angiosperms has enabled several important conclusions to be made on the likely state of the flower and carpel in the angiosperms' ancestor (Endress and Igersheim, 2000; Endress, 2001). According to these studies, the flowers of the ancestral angiosperm were probably small, bisexual, and protogynous. Its carpels were likely to have been simple (apocarpic) and incompletely closed by cellular structures, instead being sealed by substances secreted from the carpel margins. The stigmas of the angiosperms' ancestor were probably covered in muticellular protrusions and secretory. Its carpels are likely to have contained single ovules, which would probably have shown anatropous placentation, been covered by two integuments and possessed a large (crassinucellar) nucellus. It is furthermore likely that the embryo sac in the ancestral ovule was four-celled, rather than seven-celled as in most extant angiosperms (Williams and Friedman, 2002, 2004). Double fertilization would have been present in the ancestor of the angiosperms as in extant groups, leading to the production of an embryo and a biparental endosperm. However, this endosperm was most probably diploid, rather than triploid as in later-diverging groups (Williams and Friedman, 2002, 2004). Self-incompatibility (SI) systems operating between female

tissues and pollen grains are present in some ANITA angiosperms, including *Austrobaileya scandens* (Prakash and Alexander, 1984) and *Trimenia moorei* (Bernhardt *et al.*, 2003). However, it is uncertain whether homologous SI systems are to be found in any two lineages that separated at an early stage in angiosperm evolution, leaving open the question of SI as an ancestral trait in the angiosperms.

Using molecular techniques to compare ANITA angiosperms with model plants, the mechanisms likely to have controlled carpel development in the ancestral angiosperm can now be analysed. Phylogenetic analyses of the MADS box family in ANITA angiosperms and gymnosperms clearly indicate that duplication events took place in at least three MADS box lineages, the B-, C- and E-function lineages, prior to the common ancestor of the living flowering plants. These duplications may have been caused by a large-scale genomic duplication in the pre-angiosperm lineage, evidence of which is present in the *A. thaliana* genome, as reviewed by De Bodt *et al.* (2005). The pre-angiosperm C-function duplication generated two clades, respectively containing the clade-defining genes *AG* from *A. thaliana*, and *FLORAL BINDING PROTEIN7* and *11* (*FBP7/11*) from *Petunia hybrida* (reviewed by Kramer *et al.*, 2004). The *AG* clade contains angiosperm C-function genes, while the *FBP7* clade contains genes involved in ovule development in both *P. hybrida* and *A. thaliana*. The role of *FBP7*-like genes in ovule development has been defined as a new floral genetic function, the D-function (Angenent *et al.*, 1995; Colombo *et al.*, 1995), although it is not clear how widely the D-function concept applies within the flowering plants. The *FBP7* clade may have been lost from some angiosperm groups, including the Ranunculales of the basal eudicots (Kramer *et al.*, 2004).

A further duplication occurred in the ancestral E-function gene to generate two distinct E-function sub-clades in the pre-angiosperm lineage. *SEP1*, *SEP2*, and *SEP4* from *A. thaliana* appear to be descended from one of the paralogues generated by this ancient duplication, while *SEP3* appears to be descended from the other (Zahn *et al.*, 2005). As these two *SEP* sub-clades play largely redundant roles in *A. thaliana*, the functional significance of the proposed pre-angiosperm E-function duplication is not yet entirely clear.

The expression patterns of C- and E-function genes in basal angiosperms have recently been analysed (Kim *et al.*, 2005), as summarized in Fig. 2c. Expression of C-function genes is mostly limited to the third and fourth floral whorls in ANITA taxa, while E-function genes are expressed in all floral organs. These expression patterns closely resemble those of C- and E-function genes in *A. thaliana*, suggesting that important elements of the control of carpel identity may have been conserved throughout angiosperm evolution. Despite the apparent conservation of C-function expression, Kim *et al.* (2005) noted some expression of C-function genes in the perianth organs of two ANITA taxa, *Amborella* (Amborellales) and *Illicium* (Austrobaileyales). However, this observation may be related to the rather gradual transition of floral organ types that is frequently apparent in ANITA angiosperms, with intermediate forms of floral organs present at whorl boundaries (Kim *et al.*, 2005; Fig. 2c).

In addition to MADS box floral homeotic genes, the expression patterns of two further carpel development genes have recently been analysed in basal angiosperms. One of these, *CRABS CLAW* (*CRC*), encodes a transcription factor of the YABBY class. YABBY genes play roles in the specification of abaxial cellular identity of plant lateral organs by defining the side of these organs that faces away from the developmental axis (Bowman, 2000). *CRC* is expressed in the abaxial tissues of the *A. thaliana* gynoecium and in nectaries (Bowman and Smyth, 1999). A putative orthologue from the ANITA angiosperm *Amborella trichopoda* shows a similar pattern of expression in carpels to that of *CRC* from *A. thaliana* (Fourquin *et al.*, 2005), suggesting these two genes to have conserved a common developmental role since the speciation event that separated their lineages at the base of the flowering plants. Similarly, *TOUSLED* (*TSL*), encoding a serine-threonine protein kinase, shows conserved expression patterns between *A. thaliana* and the ANITA angiosperm *Cabomba aquatica* (Nymphaeales, Cabombaceae). *TSL* is necessary for normal development of the carpel apex in *A. thaliana* and shows a peak of expression in that tissue (Roe *et al.*, 1997). The orthologue of *TSL* from *C. aquatica* is also expressed at a high level in the carpel apex (Fourquin *et al.*, 2005), suggesting a conservation of function since the common ancestor of the flowering plants.

The control of carpel identity in monocots

The monocots form a monophyletic group of angiosperms whose lineage diverged later those of the ANITA grade, perhaps around 145 MYA (Davies *et al.*, 2004; Fig. 1). This group has undergone considerable evolutionary divergence to form over 60 000 extant species. Genes controlling floral organ identity have been analysed principally in two monocot models, rice and maize, both from the Poaceae or grass family. Phylogenetic analyses suggest at least one major gene duplication event to have occurred in the MADS box C-clade prior to the separation of the rice and maize lineages, with an additional subsequent duplication in one of the two sub-clades generated, specifically in the maize lineage. Accordingly, the rice C-clade gene *OsMADS58* appears orthologous to the maize gene *ZAG1*, while *OsMADS3* from rice is putatively orthologous to the two paralogous maize genes, *ZMM2* and *ZMM23* (Mena *et al*., 1996; Yamaguchi *et al.*, 2005).

Phenotypes associated with mutations in C-clade genes have been investigated in both rice and maize, although more thoroughly in the former of these species. The

inactivation of *OsMADS58* in rice leads to defects in, though does not eliminate, carpel development (Yamaguchi *et al.*, 2005; Fig. 2d). In addition, *osmads58* mutants show reduced floral determinacy, indicating a major contribution of this gene to the 'stop' function. The inactivation of *OsMADS3* eliminates stamen development, but has little or no effect on either carpel development or flower determinacy (Kang *et al.*, 1998; Yamaguchi *et al.*, 2005). Rice plants in which both *OsMADS3* and *OsMADS58* have been inactivated produce aberrant carpels, similar to those of *osmads58* single mutants, indicating *OsMADS3* to make no significant contribution to carpel development (Yamaguchi *et al.*, 2005). In maize, *zag1* mutants show a defect in floral determinacy, indicating functional conservation of *ZAG1* with its rice orthologoue *OsMADS58*. In addition, further genes that have yet to be identified are also required for female flower determinacy in maize (Laudencia-Chingcuanco and Hake, 2002).

Data from rice and maize therefore indicate the past occurrence of sub-functionalization events between two C-function gene clades in the monocots. By comparison with *A. thaliana*, a partial separation of male- and female-acting components of the C-function is apparent in grasses, with one sub-clade acting principally in stamens and the other in the fourth floral whorl to arrest organ proliferation. Interestingly, the persistence of carpel development in rice plants that lack any active C-clade MADS box genes indicates a potentially important difference in the mechanism of carpel specification between grasses and *A. thaliana*.

In contrast to the effect of inactivating C-clade MADS box genes, carpels are entirely replaced by ectopic stamens in rice plants in which the YABBY family gene *DROOPING LEAF* (*DL*) has been inactivated (Yamaguchi *et al.*, 2004; Fig. 2d). DL is also required for normal leaf development. *DL* expression has been shown to be maintained in the carpels of rice plants in which both *OsMADS3* and *OsMAD58* have been inactivated (Yamaguchi *et al.*, 2005), demonstrating its action to be independent of these. It is not yet clear whether carpel development depends on *DL* expression *per se*, or whether *DL* is mainly responsible for preventing B-function gene expression in the fourth whorl. Experiments that combine B-clade, C-clade, and *dl* mutations in rice will be needed to evaluate the relative contributions of MADS box genes and *DL* to the specification of carpel identity. *DL* appears to be orthologues to *CRC* from *A. thaliana*. The conservation of expression patterns of *CRC* orthologues between *A. thaliana* and very basal angiosperms (Fourquin *et al.*, 2005), as discussed above, suggests that the distinct roles of *DL* in carpel identity and leaf development (Yamaguchi *et al*., 2004) arose specifically in the monocot lineage.

SEP-like genes, necessary for carpel development in eudicots, are also known from monocots. *OsMADS1* from rice, corresponding to the *LEAFY HULL STERILE1* locus, groups within the same clade as *SEP1*, *2*, and *4* from *A. thaliana* (Zahn *et al*., 2005). Outer whorl floral organs in *osmads1* loss-of-function mutants take on a leaf-like appearance, whereas inner whorl floral organs are partially converted to paleas and lemmas, which are normally found in the first whorl of rice flowers (Agrawal *et al.*, 2005). These results suggest *OsMADS1* to be a principal component of the E-function in rice (Fig. 2d), while the functions of four other rice *SEP* clade genes, *OsMADS5*, *OsMADS7*, *OsMADS8*, and *RMADS217* (Zahn *et al.*, 2005), remain to be fully investigated.

Gene duplication and carpel evolution in the core eudicots

The core eudicots form a monophyletic group that is estimated to have diverged from the more basal lineages of eudicots around 110 MYA (Davies *et al.*, 2004; Fig. 1). The core eudicot clade includes all of the well-known dicot model taxa such as *Arabidopsis*, *Petunia*, and *Antirrhinum*. Analysis of the *A. thaliana* complete genome sequence has provided evidence of a large-scale duplication event that may have occurred at around the time of the ancestor of the core eudicots (De Bodt *et al.*, 2005). Evidence of this duplication can also be found in the MADS box families present in extant taxa. The comparison of diverse core eudicot groups has provided an excellent opportunity to study evolutionary events such as sub-functionalization and neo-functionalization (Moore and Purugganan, 2005), several of which are evident in eudicot genes controlling carpel, fruit, and ovule development and floral determinacy.

In the core eudicots, two gene lineages are present in place of an ancestral C-function lineage whose single descendant is present in basal eudicots. In *A. thaliana*, one of the novel lineages, the *AG* lineage, contains the *AG* gene itself, while the other, the *PLENA* (*PLE*) lineage (Fig. 3), contains a pair of paralogous genes termed *SHATTERPROOF1* and *2* (*SHP1/2*). In *Antirrhinum majus*, the probable orthologue of *AG* is termed *FARINELLI* (*FAR*), while that of *SHP1/2* is the clade-defining gene *PLENA* (*PLE*). Interestingly, the non-orthologous genes *AG* and *PLE* are responsible for specifying the C-function in *A. thaliana* and *A. majus*, respectively (Davies *et al.*, 1999; Kramer *et al.*, 2004; Fig. 3). *FAR*, by contrast, is redundantly involved in stamen development and is also required for pollen fertility in *A. majus*, while *SHP1* and *SHP2* redundantly play a novel role in *A. thaliana* fruit development (Liljegren *et al.*, 2000). In *Petunia hybrida*, which is more closely related to *Antirrhinum* than to *Arabidopsis* (Fig. 1), a further case of sub-functionalization is apparent, where the *AG* orthologue *PMADS3* is principally responsible for stamen development (Kapoor *et al.*, 2002), but probably also plays redundant roles with the *PLE* orthologue *FLORAL BINDING PROTEIN6* (*FBP6*) in both carpel development and floral determinacy (Kramer *et al.*, 2004).

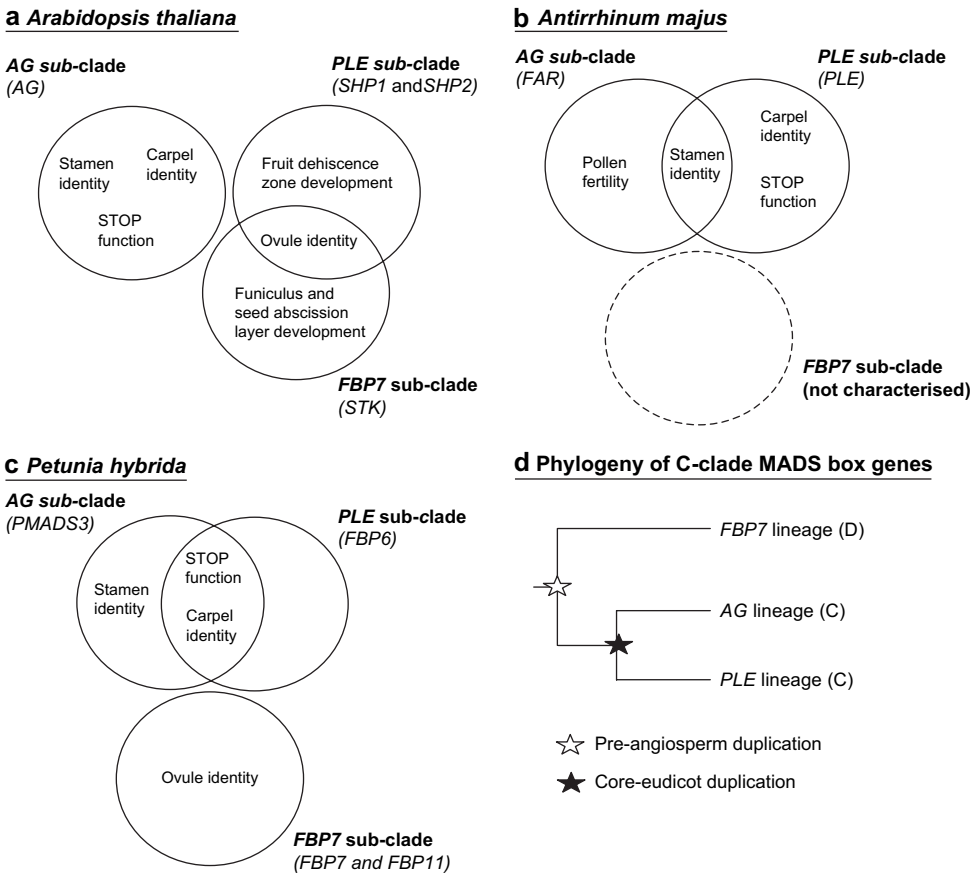


Fig. 3. Fluidity in the functionalization of C/D-clade MADS box genes in the core eudicots. (a–c) Venn diagrams representing the known functions of three C/D-function MADS box sub-clades (*AG*, *PLE*, and *FBP7*) in wild-type plants of three species from the core eudicots. Overlapping regions represent functional redundancy between genes from different sub-clades in wild-type genetic backgrounds (*AG*, *AGAMOUS*; *FAR*, *FARINELLI*; *FBP*, *FLORAL BINDING PROTEIN*; *PLE*, *PLENA*; *SHP*, *SHATTERPROOF*). (d) The phylogeny of the eudicot C/D-MADS box gene clade.

Although sub-functionalization between the paralogous *AG* and *PLE* clades in *A. thaliana* (respectively represented by the genes *AG* and *SHP1/2*) has left *AG* playing the major C-function role, elegant experiments involving multiple mutants show that the *SHP* genes have retained a capacity for C-function activity. Ectopic carpelloid organs may develop in the first floral whorl of plants lacking an active *AG* gene if the *APETALA2* (*AP2*) A-function gene is additionally inactivated (Bowman *et al.*, 1991). This effect is thought to occur because *AP2* is responsible for down-regulating C-clade genes in the outer floral whorls of wild-type plants. In the case of *ag/ap2* double mutants, the C-function activity responsible for specifying ectopic carpel development in the first whorl is provided by *SHP1* and *SHP2*, evidenced by the fact that first whorl organs of *ap2/ag/shp1/shp2* quadruple mutants are devoid of carpelloid features (Pinyopich *et al.*, 2003). These data indicate a subtle effect of functional overlap between paralogous gene clades that does not equate to simple genetic redundancy.

The fluidity of functions among duplicated genes is further illustrated by an exchange of function between C- and D-clade MADS box genes in the eudicots. Two paralogous D-function genes in *P. hybrida*, *FBP7* and *FBP11*, are redundantly essential for ovule development (Angenent *et al.*, 1995). The probable *A. thaliana* orthologue of these two genes, *SEEDSTICK* (*STK*), is also involved in ovule development, but in this case the redundancy relationship extends beyond the D-clade to include the genes *SHP1* and *SHP2* of the *PLE* sub-clade (Fig. 3). Accordingly, the *fpb7/fpb11* double mutant of *P. hybrida* (Angenent *et al.*, 1995) is phenotypically similar to the *stk/shp1/shp2* triple mutant of *A. thaliana* (Pinyopich *et al.*, 2003). Both of these mutants possess supernumerary carpels in the place of ovules within the gynoecium. In addition to its redundant role in ovule specification, *STK* plays non-redundant roles in the development of the funiculus and in seed abscission in *A. thaliana* (Pinyopich *et al.*, 2003). The C/D-function gene clade in the eudicots therefore represents a complex situation, where evolutionary processes including sub-functionalization, exchanges of function between paralogous genes, exchanges of function between non-paralogous genes, and neo-functionalization, have all taken place (Fig. 3).

The A-function gets into carpel development

A further likely consequence of the hypothesized genome duplication at the base of the core eudicots was the generation of a second sub-clade of MADS box genes within the A-function clade (Litt and Irish, 2003). The A-function MADS box gene *APETALA1* (*AP1*) plays roles in floral meristem patterning and the specification of perianth (petal and sepal) organ identity in *A. thaliana*. This latter role corresponds to the A-function, as defined by the ABCE model. However, gene (or genome) duplication in the core eudicots has provided further A-clade sequences, one of which appears to have been recruited to carpel and fruit development somewhere along the *A. thaliana* lineage. The A-clade gene *FRUITFULL* (*FUL*) is involved in the patterning of the gynoecium and fruit wall in *A. thaliana* (Gu *et al.*, 1998). *FUL* is known to act in a network involving a large number of genes (Roeder *et al.*, 2003; Liljegren *et al.*, 2004), including the MADS box genes *SHP1* and *2* (Ferrandiz *et al.*, 2000) that also function redundantly with *STK* in ovule development. Gene duplication in the A-function clade of MADS box genes, possibly caused by a whole genome duplication event, has thus resulted in novel fruit shattering mechanisms in the Brassicaceae by the process of neo-functionalization.

An interesting feature of gene-duplication in the A-clade is the evolution of a distinct C-terminal protein motif in the *AP1* sub-clade, apparently produced by a frame-shift mutation that occurred towards the 3′-extremity of the coding sequence (Litt and Irish, 2003). This frame-shift created a farnesylation site that is known to be post-translationally modified *in vivo* in *A. thaliana* and which is required for wild-type *AP1* protein activity (Yalovsky *et al.*, 2000). Other frame-shift mutations in duplicated genes are present in the B- and C-function MADS box clades of the eudicots (Vandenbussche *et al.*, 2003). However, the conserved motifs generated in these cases are distinct from that of the *AP1* lineage and do not contain farnesylation sites. The novel C-terminal motifs present in certain lineages within the eudicot A, B, and C MADS box clades have been conserved over a very long period, clearly indicating their functional significance. However, it is not yet known whether the functions of these novel motifs are connected with biochemical processes in common, such as the higher-order assembly or sequestration of MADS box transcription factor complexes (Vandenbussche *et al.*, 2003).

The carpel of the future

Many of the key questions of carpel evolution remain to be answered. For example, it is not known to which organ in gymnosperms the carpel is homologous. The mechanism of carpel closure and the potentially diverse mechanisms of fusion between carpels have yet to be discovered in the more highly evolved syncarpic species (Armbruster *et al.*, 2002). Little is known of how stigma, style, and ovary differentiation occurs in model plants, and certainly there is no information on how these processes first evolved. Although very good progress has been made to unravel the mechanisms of fruit development in *A. thaliana*, many other forms of angiosperm fruits have not yet been investigated at a molecular level.

Future research aimed at understanding carpel evolution will undoubtedly be helped by the extension of functional genetic approaches to non-model taxa. Such technological advances will depend to some extent on the development of plant transformation procedures in non-model plants, permitting the use of such techniques as RNAi (Smith *et al.*, 2000) or the directed mis-expression of transgenes. However, apart from the eudicots and monocots, many of the key taxa to be studied are of a woody habit and take several years before flowering. One technique that may help to overcome this practical difficulty is that of Virus Induced Gene Silencing (VIGS), reviewed by Burch-Smith *et al.* (2004). The VIGS technique involves the use of a transgenic virus that can inactivate a given plant gene through an RNAi-related mechanism. Several different VIGS vectors have been developed and recent studies show that at least one of these (Liu *et al.*, 2004) can infect relatively basal groups of angiosperms (Hileman *et al.*, 2005).

Acknowledgements

Recent ‘evo-devo’ research in our laboratory has been funded by a French Science Ministry grant of the ‘Action Concertée Incitative—Biologie intégrative du développement et de la physiologie’.

References

Agrawal G, Abe K, Yamazaki M, Miyao A, Hirochika H. 2005. Conservation of the E-function for floral organ identity in rice revealed by the analysis of tissue culture-induced loss-of-function mutants of the *OsMADS1* gene. *Plant Molecular Biology* **59,** 125–135.

Angenent GC, Franken J, Busscher M, Vandijken A, Vanwent JL, Dons HJM, Vantunen AJ. 1995. A novel class of MADS box genes is involved in ovule development in *Petunia*. *The Plant Cell* **7,** 1569–1582.

Aoki S, Uehara K, Imafuku M, Hasebe M, Ito M. 2004. Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *Journal of Plant Science Research* **117,** 229–244.

Armbruster WS, Debevec EM, Willson MF. 2002. Evolution of syncarpy in angiosperms: theoretical and phylogenetic analyses of the effects of carpel fusion on offspring quantity and quality. *Journal of Evolutionary Biology* **15,** 657–672.

Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, Wolfe AD, dePamphilis CW. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences, USA* **97,** 13166–13171.

Becker A, Kaufmann K, Freialdenhoven A, Vincent C, Li MA, Saedler H, Theissen G. 2002. A novel MADS-box

gene subfamily with a sister-group relationship to class B floral homeotic genes. *Molecular Genetics and Genomics* **266,** 942–950.

Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution* **29,** 464–489.

Becker A, Winter KU, Meyer B, Saedler H, Theissen G. 2000. MADS-box gene diversity in seed plants 300 million years ago. *Molecular Biology and Evolution* **17,** 1425–1434.

Bernhardt P, Sage T, Weston P, Azuma H, Lam M, Thien LB, Bruhl J. 2003. The pollination of *Trimenia moorei* (Trimeniaceae): floral volatiles, insect/wind pollen vectors and stigmatic self-incompatibility in a basal angiosperm. *Annals of Botany* **92,** 445–458.

Bowman JL. 2000. The YABBY gene family and abaxial cell fate. *Current Opinion in Plant Biology* **3,** 17–22.

Bowman JL, Smyth DR. 1999. *CRABS CLAW*, a gene that regulates carpel and nectary development in *Arabidopsis*, encodes a novel protein with zinc finger and helix–loop–helix domains. *Development* **126,** 2387–2396.

Bowman JL, Smyth DR, Meyerowitz EM. 1991. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112,** 1–20.

Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP. 2004. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant Journal* **39,** 734–746.

Carlsbecker A, Tandre K, Johanson U, Englund M, Engstrom P. 2004. The MADS-box gene *DAL1* is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *The Plant Journal* **40,** 546–557.

Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353,** 31–37.

Colombo L, Franken J, Koetje E, Vanwent J, Dons HJM, Angenent GC, Vantunen AJ. 1995. The *Petunia* MADS box gene *FBP11* determines ovule identity. *The Plant Cell* **7,** 1859–1868.

Davies B, Motte P, Keck E, Saedler H, Sommer H, Schwarz-Sommer Z. 1999. *PLENA* and *FARINELLI*: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. *EMBO Journal* **18,** 4023–4034.

Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences, USA* **101,** 1904–1909.

De Bodt S, Maere S, Van de Peer Y. 2005. Genome duplication and the origin of angiosperms. *Trends in Ecology and Evolution* **20,** 591–597.

Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. 2004. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Current Biology* **14,** 1935–1940.

Dornelas MC, Rodriguez APM. 2005. A *FLORICAULA/LEAFY* gene homolog is preferentially expressed in developing female cones of the tropical pine *Pinus caribaea* var. *caribaea*. *Genetics and Molecular Biology* **28,** 299–307.

Endress PK. 2001. The flowers in extant basal angiosperms and inferences on ancestral flowers. *International Journal of Plant Sciences* **162,** 1111–1140.

Endress PK, Igersheim A. 2000. Gynoecium structure and evolution in basal angiosperms. *International Journal of Plant Sciences* **161,** S211–S223.

Ferrandiz C, Liljegren SJ, Yanofsky MF. 2000. Negative regulation of the *SHATTERPROOF* genes by *FRUITFULL* during *Arabidopsis* fruit development. *Science* **289,** 436–438.

Fourquin C, Vinauger-Douard M, Fogliani B, Dumas C, Scutt CP. 2005. Evidence that *CRABS CLAW* and *TOUSLED* have conserved their roles in carpel development since the ancestor of the extant angiosperms. *Proceedings of the National Academy of Sciences, USA* **102,** 4649–4654.

Frohlich MW. 2003. An evolutionary scenario for the origin of flowers. *Nature Reviews Genetics* **4,** 559–566.

Frohlich MW, Parker DS. 2000. The mostly male theory of flower evolutionary origins: from genes to fossils. *Systematic Botany* **25,** 155–170.

Gomez-Mena C, de Folter S, Costa MMR, Angenent GC, Sablowski R. 2005. Transcriptional program controlled by the floral homeotic gene *AGAMOUS* during early organogenesis. *Development* **132,** 429–438.

Gu Q, Ferrandiz C, Yanofsky MF, Martienssen R. 1998. The *FRUITFULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* **125,** 1509–1517.

Hileman LC, Drea S, de Martino G, Litt A, Irish VF. 2005. Virus-induced gene silencing is an effective tool for assaying gene function in the basal eudicot species *Papaver somniferum* (opium poppy). *The Plant Journal* **44,** 334–341.

Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409,** 525–529.

Ito T, Wellmer F, Yu H, Das P, Ito N, Alves-Ferreira M, Riechmann JL, Meyerowitz EM. 2004. The homeotic protein *AGAMOUS* controls microsporogenesis by regulation of *SPOROCYTELESS*. *Nature* **430,** 356–360.

Jager M, Hassanin A, Manuel M, Le Guyader H, Deutsch J. 2003. MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. *Molecular Biology and Evolution* **20,** 842–854.

Kang HG, Jeon JS, Lee S, An GH. 1998. Identification of class B and class C floral organ identity genes from rice plants. *Plant Molecular Biology* **38,** 1021–1029.

Kapoor M, Tsuda S, Tanaka Y, Mayama T, Okuyama Y, Tsuchimoto S, Takatsuji H. 2002. Role of petunia *pMADS3* in determination of floral organ and meristem identity, as revealed by its loss of function. *The Plant Journal* **32,** 115–127.

Kaufmann K, Anfang N, Saedler H, Theissen G. 2005. Mutant analysis, protein–protein interactions and subcellular localization of the *Arabidopsis* B-sister (ABS) protein. *Molecular Genetics and Genomics* **274,** 103–118.

Kim S, Koh J, Yoo MJ, Kong HZ, Hu Y, Ma H, Soltis PS, Soltis DE. 2005. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *The Plant Journal* **43,** 724–744.

Kramer EM, Jaramillo MA, Di Stilio VS. 2004. Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* **166,** 1011–1023.

Kuzoff RK, Gasser CS. 2000. Recent progress in reconstructing angiosperm phylogeny. *Trends in Plant Science* **5,** 330–336.

Lauencia-Chingcuanco D, Hake S. 2002. The *indeterminate floral apex1* gene regulates meristem determinacy and identity in the maize inflorescence. *Development* **126,** 2629–2638.

Liljegren SJ, Ditta GS, Eshed HY, Savidge B, Bowman JL, Yanofsky MF. 2000. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404,** 766–770.

Liljegren SJ, Roeder AHK, Kempin SA, Gremski K, Ostergaard L, Guimil S, Reyes DK, Yanofsky MF. 2004. Control of fruit patterning in *Arabidopsis* by *INDEHISCENT*. *Cell* **116,** 843–853.

Litt A, Irish VF. 2003. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* **165,** 821–833.

Liu Y, Nakayama N, Schiff M, Litt A, Irish VF, Dinesh-Kumar SP. 2004. Virus induced gene silencing of a *DEFICIENS* ortholog in *Nicotiana benthamiana*. *Plant Molecular Biology* **54,** 701–711.

Mathews S, Donoghue MJ. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286,** 947–950.

Mena M, Ambrose BA, Meeley RB, Briggs SP, Yanofsky MF, Schmidt RJ. 1996. Diversification of C-function activity in maize flower development. *Science* **274,** 1537–1540.

Moore RC, Purugganan MD. 2005. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* **8,** 122–128.

Mouradov A, Glassick T, Hamdorf B, Murphy L, Fowler B, Marla S, Teasdale RD. 1998. *NEEDLY*, a *Pinus radiata* ortholog of *FLORICAULA/LEAFY* genes, expressed in both reproductive and vegetative meristems. *Proceedings of the National Academy of Sciences, USA* **95,** 6537–6542.

Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L. 2002. The *TRANSPARENT TESTA16* locus encodes the *ARABIDOPSIS B-SISTER* MADS domain protein and is required for proper development and pigmentation of the seed coat. *The Plant Cell* **14,** 2463–2479.

Parenicova L, de Folter S, Kieffer M, *et al.* 2003. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *The Plant Cell* **15,** 1538–1551.

Parkinson CL, Adams KL, Palmer JD. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* **9,** 1485–1488.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405,** 200–203.

Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF. 2003. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424,** 85–88.

Prakash N, Alexander JH. 1984. Self-incompatibility in *Austrobaileya scandens*. In: Williams EG, Knox RB, eds. *Pollination '84*. Melbourne, University of Melbourne, 214–216.

Qiu YL, Lee J, Whitlock BA, Bernasconi-Quadroni F, Dombrovska O. 2001. Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? *Molecular Biology and Evolution* **18,** 1745–1753.

Qiu YL, Lee JH, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen ZD, Savolainen V, Chase MW. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402,** 404–407.

Roe JL, Nemhauser JL, Zambryski PC. 1997. *TOUSLED* participates in apical tissue formation during gynoecium development in *Arabidopsis*. *The Plant Cell* **9,** 335–353.

Roeder AHK, Ferrandiz C, Yanofsky MF. 2003. The role of the *REPLUMLESS* homeodomain protein in Patterning the *Arabidopsis* fruit. *Current Biology* **13,** 1630–1635.

Smith NA, Singh SP, Wang MB, Stoutjesdijk PA, Green AG, Waterhouse PM. 2000. Total silencing by intron-spliced hairpin RNA. *Nature* **407,** 319–320.

Soltis PS, Soltis DE, Chase MW. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402,** 402–404.

Stellari GM, Jaramillo MA, Kramer EM. 2004. Evolution of the *APETALA3* and *PISTILLATA* lineages of MADS-box-containing genes in the basal angiosperms. *Molecular Biology and Evolution* **21,** 506–519.

Tandre K, Albert VA, Sundas A, Engstrom P. 1995. Conifer homologs to genes that control floral development in angiosperms. *Plant Molecular Biology* **27,** 69–78.

Tandre K, Svenson M, Svensson ME, Engstrom P. 1998. Conservation of gene structure and activity in the regulation of reproductive organ development of conifers and angiosperms. *The Plant Journal* **15,** 615–623.

Theissen G, Becker A. 2004. Gymnosperm orthologues of class B floral homeotic genes and their impact on understanding flower origin. *Critical Reviews in Plant Sciences* **23,** 129–148.

Theissen G, Saedler H. 2001. Plant biology: floral quartets. *Nature* **409,** 469–471.

Vandenbussche M, Theissen G, Van de Peer Y, Gerats T. 2003. Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. *Nucleic Acids Research* **31,** 4401–4409.

Williams JH, Friedman WE. 2002. Identification of diploid endosperm in an early angiosperm lineage. *Nature* **415,** 522–526.

Williams JH, Friedman WE. 2004. The four-celled female gametophyte of *Illicium* (Illiciaceae; Austrobaileyales): implications for understanding the origin and early evolution of monocots, eumagnoliids, and eudicots. *American Journal of Botany* **91,** 332–351.

Winter KU, Saedler H, Theissen G. 2002. On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an orthologue from the gymnosperm *Gnetum*. *The Plant Journal* **31,** 457–475.

Yalovsky S, Rodriguez-Concepcion M, Bracha K, Toledo-Ortiz G, Gruissem W. 2000. Prenylation of the floral transcription factor *APETALA1* modulates its function. *The Plant Cell* **12,** 1257–1266.

Yamaguchi T, Lee DY, Miyao A, Hirochika H, An G, Hirano H-Y. 2005. Functional diversification of the two C-class MADS box genes, *OsMADS3* and *OsMADS58* in *Oryza sativa*. *The Plant Cell* **18,** 15–28.

Yamaguchi T, Nagasawa N, Kawasaki S, Matsuoka M, Nagato Y, Hirano H-Y. 2004. The YABBY gene *DROOPING LEAF* regulates carpel specification and midrib development in *Oryza sativa*. *The Plant Cell* **16,** 500–509.

Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis DE, de Pamphilis CW, Ma H. 2005. The evolution of the *SEPALLATA* subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* **169,** 2209–2223.

Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ. 2002. The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences, USA* **99,** 6848–6853.

Zhang PY, Tan HTW, Pwee KH, Kumar PP. 2004. Conservation of class C function of floral organ development during 300 million years of evolution from gymnosperms to angiosperms. *The Plant Journal* **37,** 566–577.